

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**







Db 1023 TAAAGGTTGGGCTACACCTGAGAGAGTGGGTACCATTAAGAGGTGGCCAAACGTA 1082  
 Qy 1081 atgctctgagaaacatgctgagaaatcctgggtgtcaaaatctcccgagcgcaaccgtcca 1140  
 Db 1083 ATGCTCTGAGAAACATGCTGAGATCTCTGGGTTCAAAGTTCCCGAGCGGACCTGCCA 1142  
 Qy 1141 agcagactcaaatcaagaagaagctccagtaagaacccaagagacgagagagaaag 1200  
 Db 1143 ACCGACACTCAAAATCAGAGAGAACTCCAGTAAGAAACCAAGAGAGAGAGAGAAAG 1202  
 Qy 1201 taacgtttttgaaccttaacccctggagatgaaatggaactagttaacaggaagagat 1260  
 Db 1203 TAAAGTTTTTTGAACCTTACCCCTGGGATGAAATGGAATGAACTACTTAACAGAGAGAGAG 1262  
 Qy 1261 tcaagatgctcttaaccttaacatcaagcagctgcaagctggaattctcccatgtgtccg 1320  
 Db 1263 TCAGAGATGCTTATCTTAAACCATCAGACCTGCCACCTGGAATTTCTCCCATGTGCTCG 1322  
 Qy 1321 aagttgccaagcgtgtcgggttagtcaaggaacacacccaagaattcaaccaagcgag 1380  
 Db 1323 AAGTTGCCAGGCTGTGGGGTTAGTCAAGGACACCAACAAAGATTTCCACAGGGCAG 1382  
 Qy 1381 ctccaaatccctgccaaagcaagtaactgtccatgataagccgagagtgtgttacggg 1440  
 Db 1383 CTCCAATCTCTGCAAGGACGAGTAACTGATGATGATGATGATGATGATGATGATGATG 1442  
 Qy 1441 gaaccctgcgcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1500  
 Db 1443 GCACCTGGCCACAGCCGAGACCATTTTAAAGATTAACATCTCTTAAAGCCACATGACCCC 1502  
 Qy 1501 atcgacctgcacatgaacctctgagcaactgtactacattcccaagcagcagagatcc 1560  
 Db 1503 ATGGACTCTGCACATGACCTCTGACCACTGTACTACTTTCACAGCCAGAGATTC 1562  
 Qy 1561 aggttgaatataaagaatttcccaagaacaagaagaagagtgtatctctcaact 1620  
 Db 1563 AGGTGGAATCAAAAGATTTTCCCAAGAACACAGAGAGAGAGAGAGAGAGAGAGAGAG 1622  
 Qy 1621 gttcccaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1680  
 Db 1623 GCTCCTCAGAGCGCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1682  
 Qy 1681 atatgctgacatgaacatttaaaagctgtctgtaagtgtgacaacagagacagagaga 1740  
 Db 1683 ATATGCTGCTGACTGACATTTTAAAGCTGTCTGATGATGATGATGATGATGATGATG 1742  
 Qy 1741 tgcacaagaacaggaatgagcagtttcagcgtgcggagagtgctgtaacattctgcgc 1800  
 Db 1743 TGCCAAGAACAGGAATGATGACAGTTTCAGCGTGCAGAGTGTGTAACCTTTCTGCGC 1802  
 Qy 1801 aaaaaccatataaaacccaatatataacttaagaataactgttttgaataatt 1860  
 Db 1803 ACNAACCATTTATTAACCCCAACATATATCTGAAATCTGAAACTGCTTGAATAATTT 1862  
 Qy 1861 ggaatactgataaactccagtgagccaagacatgtgtgataaanaagtgccaagaagcag 1920  
 Db 1863 GGAATATCTGATTAATCTCAGTGGGCCAAGACATGTGATTAATAAAGTGTGCAAGAGCAG 1922  
 Qy 1921 aagaaactcaggtgtagccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1980  
 Db 1923 AAGAAACTTCAGGTGTGATCCCTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1982  
 Qy 1981 catcaatcaagaagaacagacccaagcgtccagctcagctcagctcagctcagctcagctc 2040  
 Db 1983 CATCAATCCAGACAGAAACACCCCAAGCCCTCCAGTTCTGTTTTGATCCTGACAAA 2042  
 Qy 2041 gagagacagcacaatctcatgt 2100  
 Db 2043 GAGAGCAGAGCCATTTCTATGCTGCTTCTTCAGATACCTTTGAAAAAACCCGAGAGCCA 2102  
 Qy 2101 caacgaagagccttataagcggcccgagcctaagcagcagagagaaagcaggtgtctt 2160

Db 2103 CACCAAGAGAGCCCTTATAGCGCCCGGAGACTTAACGAGACAGAAAGGCCATGTGCTT 2162  
 Qy 2161 cctactgacatgaactgaactgaactgaactgaactgaactgaactgaactgaactgaactga 2220  
 Db 2163 CCTACTGACATGACGACTGACCTCGCCGACACAGTAGACACGACTGATACCACTGCTTCC 2222  
 Qy 2221 tcttc-agttcaatcttctccttgaattgaatacaactataatltcaatcagttcc 2279  
 Db 2223 TCTTCAAGTTTCAATTTTTTCTTCTTATGATGATCAACACACTATTAATTTTCAATTTCC 2282  
 Qy 2280 ttaagtgtctacttaacctgaagatgttaagaacttcaagtaatgaactgaagagatc 2339  
 Db 2283 TTAGTGTGTCTACTTACCTGACGATTTGAAACTGTCACTGATGATGATGATGATGATG 2342  
 Qy 2340 acagcccggttgggtgtgacatctgtccctgtgctgtgctgtgctgtgctgtgctgtgct 2399  
 Db 2343 ACAGCCCGGTGGGTGGGATCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 2402  
 Qy 2400 ccagaatccttgaactcctgtccctgtgataaatlgyacagcttaagactttaaactt 2459  
 Db 2403 CCAGATTCCTGTGACTCCGCCCCCTGTATTAATGAGAGCTTAGGACTTTAAACTTT 2462  
 Qy 2460 agatcaaaagatagtgctctttaaacttattttaaagagagagactttaaagagcc 2519  
 Db 2463 AGATCAAAAGATATGCTCTTTTAACTTTATTTTAAAGAGAGACACTTTAAATGAGCC 2522  
 Qy 2520 ctgaccttaaccattataaagaatttgcataaagaagatgttcttgaagagatagctt 2579  
 Db 2523 CTGACCTTTAACCATTTATACAGAAATTTGTCAAAAGAGATGTTTCTTGAGAGAGTAC 2582  
 Qy 2580 tttttaccacacatgaagacatlaacttgaagcagcagagactaagctgtgtcctc 2639  
 Db 2583 TTTTTCACACACTACAGGACATTTACTGTAGGCCACAGAAAGACTTACAGGCTGTGCCCT 2642  
 Qy 2640 agagggccaatacagtcgaatctcaactcctaagtcggggaaagtgacaggttccctgt 2699  
 Db 2643 AGAGGGCCCAATACAGTCAATTTCAACCTTAAGTGGGGAAAGGTGACAGGTTTCTGTG 2702  
 Qy 2700 tgcgtgtgtgcaagggagagagagtcagctgtgctgtgagagagacatgtgtgtccta 2759  
 Db 2703 TGCTGTGTGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2762  
 Qy 2760 gtgcagccctgtctcactcttgttgaactgtgaaccttccactcactgaatataagta 2819  
 Db 2763 GTGCACCCCTGCTTCACGCTTGTGTTTACTGTGAGACCTTCCACCTGATGAATATAAGTA 2822  
 Qy 2820 aactcaattcttctgcaccaataatgtaataactaa 2857  
 Db 2823 AACTCACTTCTTGTGACCAATTAATGTTAATACTAA 2860  
 RESULT 2  
 AF290989 3041 bp mRNA ROD 09-AUG-2000  
 LOCUS Rattus norvegicus Stau-16 (Stau) mRNA, complete  
 DEFINITION Rattus norvegicus Stau-16 (Stau) mRNA, complete  
 cds, alternatively spliced.  
 ACCESSION AF290989  
 VERSION AF290989.1 GI:9755162  
 KEYWORDS  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 REFERENCE 1 (bases 1 to 3041)  
 Monshausen, M., Putz, U., Reibeln, M., Schweizer, M.,  
 Desgroselliers, L., Kuhl, D., Richter, D. and Kindler, S.  
 TITLE Two rat brain stau-16 isoforms differentially bind RNA  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 3041)  
 Monshausen, M. and Kindler, S.  
 AUTHORS  
 TITLE Direct Submission  
 JOURNAL Submitted (28-JUL-2000) Institute for Cell Biochemistry and

22	122.8	4.3	2008	34	AK023314	Homo sapi
23	120.4	4.2	160979	52	AC055879	Homo sapi
24	113.2	4.0	332811	72	AL390198	Homo sapi
25	95.8	3.4	5237	53	DROSTAFEN	M69111 D.melanog
26	95.8	3.4	61267	28	AC004336	Drosophila
27	95.8	3.4	120625	38	AC017563	Drosophila
28	95.8	3.4	267271	28	AE003800	Drosophila
29	95.6	3.3	8002	29	AF225925	Musca dom
30	91.2	3.2	8409	29	AF225924	Drosophila
31	87	3.0	171591	50	AC027018	Homo sapi
32	70.6	2.5	2847	69	HSRNTTFT	X83973 H.sapiens
33	65	2.3	1166	69	AL078636	Homo sapi
34	64.2	2.2	7218	81	166494	Sequence 14
35	64	2.2	7218	81	166494	Sequence 14
36	57.8	2.0	42099	29	CELE55A4	U67949 Caenorhadi
37	54.6	1.9	212493	62	AC080167	Mus muscu
38	53.8	1.9	93821	46	AC021596	Homo sapi
39	52.8	1.8	87120	38	AC012225	Homo sapi
40	52.6	1.8	78690	38	AC016541	Homo sapi
41	52.6	1.8	164520	39	AC020738	Homo sapi
42	52.2	1.8	61363	54	AC068364	Homo sapi
43	52.2	1.8	280047	54	AC068494	Mus muscu
44	52	1.8	58514	62	AC079898	Homo sapi
45	52	1.8	64272	47	AC023019	Homo sapi

## ALIGNMENTS

RESULT 1	LOCUS	AF061942	2860 bp	mRNA	ROD	04-MAR-1999
DEFINITION	Mus musculus RNA-binding protein stauflen mRNA, complete cds.					
ACCESSION	AF061942					
VERSION	AF061942.1					
KEYWORDS	GI:4335944					
SOURCE	house mouse.					
ORGANISM	Mus musculus					

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 2860)  
Wickham, L., Duchaine, T., Luo, M., Nabi, I.R. and DesGroselliers, L.  
Mammalian stauflen is a double-stranded-RNA- and tubulin-binding protein which localizes to the rough endoplasmic reticulum  
Mol. Cell Biol. 19 (3), 2220-2230 (1999)

JOURNAL MEDLINE  
99147057  
2 (bases 1 to 2860)  
Duchaine, T., Luo, M. and DesGroselliers, L.  
Direct Submission

AUTHORS  
Submitted (28-APR-1998) Biochemistry, University of Montreal, 2900  
Edouard Montpetit, Montreal, Que H3T 1J4, Canada

FEATURES  
Source  
Location/Qualifiers

1..2860  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
326..1789  
/codon\_start=1  
/product="RNA-binding protein stauflen"  
/protein\_id="AAI17529.1"  
/db\_xref="GI:4335945"

/translation="MYKPVDPHSMQSTYSYGMNGAYPPRYFPYPPVLPVYVELS  
VGGOFGKGMKRPVHDAPARALRTLOSEPLERLVNGREAEENLNSLSOYF  
EIALKRLPVNFVARESGPRHMKNFYTRVSVGFVGBGSKSKISKKNARVLEO  
LRRLPPIVAYERYKPKIKKSOPTCKTAPDYGGMNISRLOAOQAKKEPEPYML  
TERGLPRRREVMOVKVGHHTABGVGNKVAKNAAENMLTILGEFVPOQPKPML  
KSEKTPVKRGDGRKVTFFPSPDENGNSNKEDEMRMPLSHQVPAQILPPVPEV  
AOAVGSGHHTKDFTRAAPPAKATVTAMARLLAGTSPETETILKSSISSGHVP  
HGPTRESEOLYTSRAQGVQVEXKDPKNNKNCVSLINCSOPPLVSHIGDYES  
CHDMAAIIITKLISELDQOSTEMPTRGNGVPSACGRC"

BASE COUNT  
ORIGIN

747 a 750 c 713 g 650 t

Query Match	99.6%	Score 2846	DB 74	Length 2860
Best Local Similarity	100.0%	Pred. No. 0		
Matches 2857	Conservative 0	Mismatches 0	Indels 1	Gaps 1
Qy	1	cgggcgggccccccccccgctcaactccctgcagagctgaggcccgagccgctctcag	60	
Db	3	CGGCGGGGGCCCCCCCCCGTCACTTCCTGCAAGCTGCGGGCCCCGAGCGGCTTACG	62	
Qy	61	cggttgagctgagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	120	
Db	63	CGTTTGGCGTGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG	122	
Qy	121	gcggcgcccgccctgcgctccgcaactccgctctccctctctctgctgctgctctt	180	
Db	123	CGGGCGCCCGGCTTCCGCTCCGCGCACCTCCGCTTCCCTCTTGCGGGCCCTTTT	182	
Qy	181	ccctgcgctctcaacttgctctctcaactctcgcgcgcgcgcgcgcgcgcgcgcgcg	240	
Db	183	CCTGCGGCTTCACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	242	
Qy	241	gacgagctctggggaagcagcagcagcagcagcagcagcagcagcagcagcagcagc	300	
Db	243	GACGAGCTCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	302	
Qy	301	gtctgaaacttgaaagaaaccaaactgtaaaagccgttgaccctcaactctcgatgca	360	
Db	303	GTGTGAACCTGGAAGAAACCAATGTATAGCCCGTGGAGCCCTCCTGATGAGT	362	
Qy	361	ccacctacagctatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcat	420	
Db	363	CCACCTACAGCTATGAGCATGCTGAGAGGCTGCTTACCCCGAGATACCTTATACCAT	422	
Qy	421	caagcccaactctcaactcaactcaactcaactcaactcaactcaactcaactcaact	480	
Db	423	CAGTCCACCTTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT	482	
Qy	481	aaggaagatgagacacacccgctgaacacagctgcccctgctgctgctgctgctgct	540	
Db	483	AAGGAAGATGAGACACCCGCTGAACACGATGACGATGACGATGACGATGACGATG	542	
Qy	541	agagtgaaacccctgcagaaagagtgagtgagtgagtgagtgagtgagtgagtgag	600	
Db	543	AGAGTGAACCCCTGCGACAAAGGTTGAGTGAATGGAAGGAGGAGGAGGAGGAGG	602	
Qy	601	tcaataatcgaaataaagcgaagtggttgaaattgcaattgcaattgcaattgcaatt	660	
Db	603	TCAATAATCGGAATAAGCCAAAGTGTGAAATGCGCTGAAGGGAGATTTGCTGTGA	662	
Qy	661	atttgagtgcccgagagtgagtgccacacacatgaagaacttctgacagaggtt	720	
Db	663	ATTTTGAGCTGGCCCGGAGAGTGCGCCACACATGAACATTTGTGACAGGGGTTT	722	
Qy	721	cagttgaggaaatttgtagggaaggaaggaaggaaggaaggaaggaaggaaggaatg	780	
Db	723	CAGTTGGGGAATTTGTAGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	782	
Qy	781	cggcgaagggctgctctgtagcagctttagagagcttgcacccctctctctcttgcag	840	
Db	783	CGGCGAGGGCTGTTCTTGAAGAGCTTGAAGAGCTGCCACCCCTCTCTGTGACGAG	842	
Qy	841	tgaagcccaagaaacaaagaaagaaagaaagaaagaaagaaagaaagaaagaaag	900	
Db	843	TGAAGCCCAAGATCAAGAAAGAAAGTACGCCACCTGCAAGACGCCCGGATATG	902	
Qy	901	aaggaatgaatccctatctagtgacttgcacagatccagcagcagcagcagcagcag	960	
Db	903	AAGGATGAATCCATATGAGTACTTGCACAGATCCACAGCAGCAAAAAAGAGAGG	962	
Qy	961	cagagtaacatctctctctcaagaagaggtcttcaacgtcgcgaggaaggttctgag	1020	
Db	963	CAGAGTACATGCTCTTCAAGAGAGAGGCTTCCAGCTGCGAGGAGGAGGAGGAGG	1022	
Qy	1021	taaggttgagcatcacactgcagaaagagtggttaccataaagaagtgagcgaagc	1080	

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2001, 09:22:49 ; Search time 24649.1 Seconds  
(without alignments)  
593.183 Million cell updates/sec

Title: us-09-316-048-9

Perfect score: 2857  
Sequence: 1 cggcgcgcgccctcccccgcg.....ccaataatgtaactactaa 2857

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 111813 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: GenEmbl.\*
- 2: gb\_da1.\*
- 3: gb\_da2.\*
- 4: gb\_da3.\*
- 5: gb\_da4.\*
- 6: gb\_da5.\*
- 7: gb\_da6.\*
- 8: gb\_da7.\*
- 9: gb\_da8.\*
- 10: gb\_da9.\*
- 11: gb\_da10.\*
- 12: gb\_da11.\*
- 13: gb\_da12.\*
- 14: gb\_da13.\*
- 15: gb\_da14.\*
- 16: gb\_da15.\*
- 17: gb\_da16.\*
- 18: gb\_da17.\*
- 19: gb\_da18.\*
- 20: gb\_da19.\*
- 21: gb\_da20.\*
- 22: gb\_da21.\*
- 23: gb\_da22.\*
- 24: gb\_da23.\*
- 25: gb\_da24.\*
- 26: gb\_da25.\*
- 27: gb\_da26.\*
- 28: gb\_da27.\*
- 29: gb\_da28.\*
- 30: gb\_da29.\*
- 31: gb\_da30.\*
- 32: gb\_da31.\*
- 33: gb\_da32.\*
- 34: gb\_da33.\*
- 35: gb\_da34.\*
- 36: gb\_da35.\*
- 37: gb\_da36.\*
- 38: gb\_da37.\*
- 39: gb\_da38.\*
- 40: gb\_da39.\*
- 41: gb\_da40.\*
- 42: gb\_da41.\*
- 43: gb\_da42.\*

- 44: gb\_da43.\*
- 45: gb\_da44.\*
- 46: gb\_da45.\*
- 47: gb\_da46.\*
- 48: gb\_da47.\*
- 49: gb\_da48.\*
- 50: gb\_da49.\*
- 51: gb\_da50.\*
- 52: gb\_da51.\*
- 53: gb\_da52.\*
- 54: gb\_da53.\*
- 55: gb\_da54.\*
- 56: gb\_da55.\*
- 57: gb\_da56.\*
- 58: gb\_da57.\*
- 59: gb\_da58.\*
- 60: gb\_da59.\*
- 61: gb\_da60.\*
- 62: gb\_da61.\*
- 63: gb\_da62.\*
- 64: gb\_da63.\*
- 65: gb\_da64.\*
- 66: gb\_da65.\*
- 67: gb\_da66.\*
- 68: gb\_da67.\*
- 69: gb\_da68.\*
- 70: gb\_da69.\*
- 71: gb\_da70.\*
- 72: gb\_da71.\*
- 73: gb\_da72.\*
- 74: gb\_da73.\*
- 75: gb\_da74.\*
- 76: gb\_da75.\*
- 77: gb\_da76.\*
- 78: gb\_da77.\*
- 79: gb\_da78.\*
- 80: gb\_da79.\*
- 81: gb\_da80.\*
- 82: gb\_da81.\*
- 83: gb\_da82.\*
- 84: gb\_da83.\*
- 85: gb\_da84.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2846	99.6	2860	74	AF061942 Mus muscu
2	2013.4	70.5	3041	74	AF290989 Rattus no
3	1985.4	69.5	3059	74	AF227200 Rattus no
4	1354.6	47.4	3142	10	AF061940 Homo sapi
5	1269.8	44.4	3217	10	AF061938 Homo sapi
6	1256	44.0	3506	10	AF061939 Homo sapi
7	1255	43.9	3349	10	AF061941 Homo sapi
8	1253	43.9	3194	34	AK025519 Homo sapi
9	1209.2	42.3	3066	68	HSB132258 Homo sapi
10	1045.2	36.6	225016	52	AC063967 Mus muscu
11	325.4	11.4	360	74	RN0010200 Rattus no
12	283.6	9.9	113836	34	AL133174 Human DNA
13	279.4	9.8	2968	34	AK002152 Homo sapi
14	279.4	9.8	4058	68	HSB14062 Homo sapien
15	210.2	7.4	163253	68	HSB140E4 Human DNA
16	188.4	6.6	7241	74	RN0010200 Rattus no
17	159	5.6	160979	52	AC055879 Homo sapi
18	158.2	5.5	194407	54	AC068845 Homo sapi
19	156.6	5.5	157803	35	AC011492 Homo sapi
20	154.6	5.4	3173	34	AK001576 Homo sapi
21	127.4	4.5	1300	74	RATCB09K Rat calbind



```
Db 205 TGTAAATTACATCTCCATTAGGGCTGAAAAAATGACCTACGTTTCTGTATACAGCTGTG 146
Qy 3151 ttgcttttgaatgtgtgttactgttacacagaagtgtgtgcacgtgagctctgtgtgt 3210
Db 145 TTGCTTTTGTATGTTGTTTACTGTACACAGAACTGTGTGACAGAGGCTGTGCTGTGCT 86
Qy 3211 ccgtatgaaacacgtgtgtagccctgcgagtttaagtaactgtcttccatcatctgtttacgc 3270
Db 85 CCTATGGAAGAGCCTGTGACCCCTGCGAGTTAAGTACTGCTCCATTCATTTGTTACGCT 26
Qy 3271 ggaatttctccacaggaatgta 3295
Db 25 GGAATTTTCTCCCATGGAATGTA 1

RESULT 15
LOCUS RNO010200 360 bp mRNA ROD 22-JAN-1999
DEFINITION Rattus norvegicus mRNA for Staufen protein, partial.
ACCESSION AJ010200
VERSION AJ010200.1 GI:4138427
KEYWORDS staufen gene.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 360)
AUTHORS Kiebler M.A.
JOURNAL Direct Submission
Submitted (16-AUG-1998) Kiebler M.A., Cell Biology, EMBL,
Meyerhofstr. 1, D-69117 Heidelberg, GERMANY
2 (bases 1 to 360)
REFERENCE Kiebler M.A., Hemraj, I., Verkade, P., Kohrmann, M., Fortes, P.,
AUTHORS Marion, R.M., Ortin, J. and Dotli, C.G.
TITLE The mammalian staufen protein localizes to the somatodendritic
domain of cultured hippocampal neurons: implications for its
involvement in mRNA transport
JOURNAL J. Neurosci. 19 (1), 288-297 (1999)
MEDLINE 99088098
FEATURES
source location/Qualifiers
1..360
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone_lib="Rat brain"
/dev_stage="embryonic"
1..119
/gene="staufen"
<1..>119
/gene="staufen"
/evidence="experimental"
/evidence_start=2
/product="Staufen Protein"
/protein_id="CAA09037.1"
/db_xref="GI:4138428"
/translation="YFPPVPLLYVELSVGGQDFNGKMRPPVKHDARA"
BASE COUNT 109 a 70 c 105 g 76 t
ORIGIN
```

Query Match 8.6%; Score 287; DB 74; Length 360;  
Best Local Similarity 87.5%; Pred. No. 2.1e-57;  
Matches 314; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

```
Qy 596 tactttaccattccagttccatttcaattatcaagtgaacttctgtggaaga 655
Db 2 TACTTTTACCATTCCAGTTCCAGTTCACTGTACCAAGTTGAGCTCTCCGTGGAGGA 61
Qy 656 cagcaattcaatggaagaagaacagagcgtgcgaagaacatgctgtgtccaa 715
Db 62 CAGCACTTTAATGGAAGGAAGGAAAGATGAGACCACTGTGAACATGATGCCACTGCCGA 121
Qy 716 gcgttgagagatccctgagaatgagccctgcagagaagcgtggaagtgaatgaaagaa 775
Db 716 gcgttgagagatccctgagaatgagccctgcagagaagcgtggaagtgaatgaaagaa 775
```

```
Db 122 GCCCTGAGACTCTGCAGAGTGAAGCCCTTACCAAGAGGCTAGAGTAAATGAGAGAA 181
Qy 776 tccgaagaagaataatcctcaataaactgaataagtglttgagaltgcactaa 835
Db 182 TCAGAAGAAACAAACCTTATTAATCAGAAATTAAGCAAGTGTGTGATGGCTGAAG 241
Qy 836 cgaacttgccgtgaaattcgaagtggccgggaagatggccaccacatgaagaac 895
Db 242 CGGAATTTGCTGTGAATTTTGAGGTGGCCCGGAGAGTGGCCGCCACATGAAGAAC 301
Qy 896 ttgtgacaaagglttcggttggaagtgtgtgtggaagtgaagtgaagaagaaga 954
Db 302 TTGTGACCAAGGTTTCAGTTGGGAGATTGTAGGGGAAGAGAGAGAAAGCAAGAA 360
```

Search completed: April 5, 2001, 09:22:49  
Job time: 80614 sec

Best Local Similarity 73.2%; Pred. No. 9.7e-80;  
Matches 586; Conservative 0; Mismatches 203; Indels 11; Gaps 7;

```

Oy 2308 gtcttttttccctgtgtgtaagaagaacgacgacccctctcaagctgtctaac 2367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60919 GTATTTTATTTGGCTGTGTAAGAAAAAGAAATGGAAGACAGAAACCATTTACCTCTCAC 60860

Oy 2368 tcaagacattggagacaacccttgacagccatgacagagaagacgttcaacggccc 2427
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60859 ACTGGCTACCTCAAC -ACCTTGCATTAACATGCTGGAGAAAGCCTTAGCTG60801

Oy 2428 agaagctaaagacacacagagaatcaaatgtcttctactacgcygacccaatttcta 2487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60800 AAAGCTGAAAACACCGGGGAAATTAACCTGCTCTCTACTCGTTACACATTACCTTCCTA 60741

Oy 2488 gtgtgccacggcccaacacacccctccgagtaaccacacacacacacacgtctctcc 2547
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60740 GCATGCTACGG -CCAGTCACTTACTGACAGTACCCACACCATTAACCATGCTTCTCTCT 60682

Oy 2548 aacagctgactgtaactctcaatctattcttcttctgattgatacactataaa 2607
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60661 AATAGTGATACATATTTTATGTTTATTA -TTTCTTTGATTGAATGATGCCATATGAA 60623

Oy 2608 atttcattgagaattcccaattgtatctagttaaaagacacagcttggaaactgtc 2667
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60622 ATTTCATTTGAAAGTTCTCTGAGTTGATGATACATAGTACAGTTCAGAAACCTGTG 60563

Oy 2668 tgaagctgacttatacaataatctaacgcgaagaatatacatcatgtgtgtgttag 2727
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60562 CCAGCTACCTTATACAGAAATCTTAACGTCATGATATATCATATACGATGATGTTGT 60503

Oy 2728 acaattctattctgactgactaacccagacagcttcaagtgatgaactgtgtgcccct 2787
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60502 ACATTCCTTTCTACTGACTACCAAGAACCATTTCACTGATGATGCAATATGATGACTCT 60443

Oy 2788 ggttcagctgaaacagctccctgactt-tcaaaaacctggaataagctccacagctgta 2846
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60442 GGTTCAGCTGAAAAAGCCCTGAGCTTCACAGAAACCTTGATCAAAATTTCCACTATGTA 60383

Oy 2847 taacttgacaatttaagaatttaactttagatgacttctgttccatttatttctc 2906
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60382 CAAATTTGCCATTTTGAAGAAATTCAACTGCA ---ATCATTTGGTCTCTTTATATTTT 60327

Oy 2907 atttattcttgtttaaagca--aacaggaacttaaatgaacttatactctgtttaag 2964
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60326 ATTTCATTTTGGTGTGTGATGAGGACCTTAACAAATAAATCTTGATCTTTTAAAG 60267

Oy 2965 attatataaacaacattgtgtatctatacatatgctcttgaggacttagcttcaactaa 3024
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60266 ATTATTAATAAACTTACGTATGTACATATAGCTCTCAAGACCCACAGCTTTCACATAG 60207

Oy 3025 ctacaggaatagatctcatatgagccatataaactctgagagtgatcttcc-aagagtc 3083
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60206 CTGTAGGATATTTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 60147

Oy 3084 tcgatactgttaataatcac 3103
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60146 TTTATACCTTATTTTTC 60127
  
```

RESULT 14  
LOCUS G30939 385 bp DNA STS 29-SEP-1998  
DEFINITION DRESS50 Human DRES Homo sapiens STS genomic clone 22368 3', sequence tagged site.  
ACCESSION G30939  
VERSION G30939.1 GI:1724032  
KEYWORDS STS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 385)

AUTHORS Banfi,S., Borsani,G., Rossi,E., Bernard,L., Guffanti,A.,  
Rubboli,F., Marchitelli,A., Giglio,S., Coluccia,E., Zollo,M.,  
Zuffardi,O. and Ballabio,A.  
IDENTIFICATION and mapping human cDNAs homologous to Drosophila  
mutant genes through EST database searching  
Nat. Genet. 13 (2), 167-174 (1996)  
96225443

CONTACT: Ballabio Andrea  
Telethon Institute of Genetics and Medicine, TIGEM  
Telethon Institute of Genetics and Medicine, TIGEM  
via Olgettina 58, 20132 Milano, Italy  
Tel: +39-2-215601  
Fax: +39-2-21560220  
Email: dres@tigem.it  
PRIMER A: GTAAACATGATGAGACGAG  
PRIMER B: GGCTGAAAAAGATGACCTAC  
STS size: 157  
PCR profile:

Protocol:  
Template: 25 ng  
Primer: each 1 uM  
dNTPs: each 220 uM  
Tag Polymerase: 0.025 units/uL  
Total Vol: 15 uL  
Presoak: 94 degrees C for 2 minutes  
Denaturation: 94 degrees C for 45 seconds  
Annealing: 57 degrees C for 45 seconds  
Polymerization: 72 degrees C for 45 seconds  
PCR cycles: 35  
Thermal Cycler: MJR PTC-200

Buffer:  
MgCl2: 1.5 mM  
KCl: 50 mM  
Tris-HCl: 10 mM  
pH: 8.3

FEATURES  
source  
1..385  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="20, 12.10 cR from GC110F11 (lod >3.0)"  
/clone="22368"  
/clone\_11b="Human DRES"  
STS  
primer\_bind  
29..49  
complement(165..185)  
BASE COUNT 135 a 79 c 71 g 100 t  
ORIGIN

Query Match 11.4%; Score 380.2; DB 75; Length 385;  
Best Local Similarity 99.2%; Pred. No. 2.5e-79;  
Matches 382; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

Oy 2911 ttatttgttaatgaacagagacttaatgaacttgatctcgttttaagattatt 2970
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 385 TTTATTTTCTTAATCAACAGAGCTTAATGAACCTTGATCTGTTTAAAGATTATT 326

Oy 2971 aaaaacattgtatctatatacatatgctcttgaggacttagcttcaactacatag 3030
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 325 AAAAAACATGTGTATCATATATAGCTCTTGAAGACTTATGCTTTCATCACTACAG 266

Oy 3031 gataatccatgtagtccatataaaccctgacagagatcttccagaagtcgtcatac 3090
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 265 GATATGATCTCCATGTAGCCCATATAAACCTGCAAGAGTATTTCCAGAGTCTGATAC 206

Oy 3091 tgttaatacatctccattagggctgaaagaatgacctgaagcttcgtatatacagctgtg 3150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  
```

Oy	2182	tgatggcaaccggttgacctggcggtccctcagcaatagcagctgtytgggaaagaag	2241
Db	3006	tgacggccactctgggagacctggctctccctacgcattccagcttctttgggagagag	2947
Oy	2242	ggtcgccagccagctggtctctccggygacacagcagatcacaccctgggcactcgy	2301
Db	2946	ggcttcccccagccagctggctctccctgggggacacagcatctccagccctgggaccttcc	2887
Oy	2302	tgattggacttttttttccctctgfg--tgaaaagaagaagcagaccccttcaagc	2359
Db	2886	tgttttggcttttttttttccctctgtaaaagaagaaactgaaagcatgaccccttctcaagc	2827
Oy	2360	tgactactcagacacatcttggaaaaaacctggagcagcatgycagaagaagagccttga	2419
Db	2826	tgggctta-----gttgggagacaaaccttgagacgacacaccagagagagccttcaa	2776
Oy	2420	ccggccccagagctataaaagcaccagagaagaataaagtcttctactcagcgtyaaccaa	2479
Db	2775	ccagccccagagctataaaagcaccagagaagaataaagtcttctactcagctaaacctaa	2716
Oy	2480	ctttctcagtagtgccagcgcccacacacctctctgagtagcccaacaacatgaacagtgtt	2539
Db	2715	c--ttctagtgtagccacgcccacacacctctctgagtagcccaacaacatgaacagtgtt	2658
Oy	2540	tctcttcacaagatgactgtatcttacttagttlcaatlacttcttttgatgataygac	2599
Db	2657	tctcttcccaacagatgactgtatcttacttagtttcaattatttctttttttatgtgatgacac	2598
Oy	2600	tatatataatttcat 2615	
Db	2597	tatatataatttctct 2582	

```

RESULT 13
AL158850/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
-----
AL158850 192597 bp DNA HTG 06-SEP-2000
Homo sapiens chromosome 6 clone RP3-522B19, *** SEQUENCING IN
PROGRAMS ***, 11 unordered pieces.
AL158850
AL158850.4 GI:9800143
HTG: HTGS_PHASE1; HTGS_DRAFT.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 192597)
Burton.J.
Direct Submission
Submitted (05-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requests: clonerequests@sanger.ac.uk
On Aug 19, 2000 this sequence version replaced gi:9212635.
-----
Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
Project Information
Center project name: d1522B19
-----
Summary Statistics
Assembly program: XGAP; version 4.5
Sequencing vector: M13; M77815; 12% of reads
Sequencing vector: plasmid; L08752; 87% of reads
Chemistry: Dye-terminator Big Dye; 77% of reads
Chemistry: Dye-primer-amersham; 22% of reads
Chemistry: Dye-primer Big Dye; 0% of reads
Consensus quality: 185625 bases at least Q40
Consensus quality: 188533 bases at least Q30
Consensus quality: 190005 bases at least Q20
Insert size: 191597; sum-of-contigs
Insert size: 144116; 9.6% error; agarose-tp
Quality coverage: 5.71x In Q20 bases; sum-of-contigs Quality

```

```

-----
* coverage: 8.21x in Q20 bases; agscore=fp
*
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 13837: contig of 13837 bp in length
* 13838 13937: gap of 100 bp
* 13938 17476: contig of 3539 bp in length
* 17477 17576: gap of 100 bp
* 17577 20005: contig of 2429 bp in length
* 20006 20105: gap of 100 bp
* 20106 32070: contig of 11965 bp in length
* 32071 32170: gap of 100 bp
* 32171 39772: contig of 7602 bp in length
* 39773 39872: gap of 100 bp
* 39873 43471: contig of 3599 bp in length
* 43472 43571: gap of 100 bp
* 43572 51795: contig of 8224 bp in length
* 51796 51895: gap of 100 bp
* 51896 178838: contig of 126943 bp in length
* 178839 178938: gap of 100 bp
* 178939 181919: contig of 2981 bp in length
* 181920 182015: gap of 100 bp
* 182020 184231: contig of 2212 bp in length
* 184232 184331: gap of 100 bp
* 184332 192597: contig of 8266 bp in length.
*
* Location/Qualifiers
*   1..192597
*     /organism="Homo sapiens"
*     /db_xref="taxon:9606"
*     /chromosome="6"
*     /clone="RPJ-522B19"
*     /clone_lib="RPCR-3"
*
* misc_feature
*   1..13837
*     /note="assembly-fragment:02383
*     fragment_chain:1"
*
* misc_feature
*   13938..17476
*     /note="assembly-fragment:01080
*     fragment_chain:1"
*
* misc_feature
*   17577..20005
*     /note="assembly-fragment:02434
*     fragment_chain:1"
*
* misc_feature
*   20106..32070
*     /note="assembly-fragment:03060
*     fragment_chain:2"
*
* misc_feature
*   32171..39772
*     /note="assembly-fragment:00945
*     fragment_chain:2"
*
* misc_feature
*   39873..43471
*     /note="assembly-fragment:00013
*     fragment_chain:2"
*
* misc_feature
*   43572..51795
*     /note="assembly-fragment:02225
*     fragment_chain:3"
*
* misc_feature
*   51896..178838
*     /note="assembly-fragment:01767
*     fragment_chain:3"
*
* misc_feature
*   178939..181919
*     /note="assembly-fragment:01637"
*     fragment_chain:3"
*
* misc_feature
*   182020..184231
*     /note="assembly-fragment:02408"
*     fragment_chain:3"
*
* misc_feature
*   184332..192597
*     /note="assembly-fragment:02960"
*
*
* BASE COUNT      58324 a 39515 c 38313 g 55451 t 1014 others
*
* ORIGIN
*
Query Match      11.4%: Score 383.2: DB 70: Length 192597.

```



```
* as soon as it is available and the accession number will
* be preserved.
1 1756 1855: contig of 1755 bp in length
* 1756 1855: gap of 100 bp
* 1856 3429: contig of 1574 bp in length
* 3430 3529: gap of 100 bp
* 3530 5346: contig of 1817 bp in length
* 5347 5446: gap of 100 bp
* 5447 7338: contig of 1892 bp in length
* 7339 7438: gap of 100 bp
* 7439 10297: contig of 2859 bp in length
* 10298 10397: gap of 100 bp
* 10398 13590: contig of 3193 bp in length
* 13591 13690: gap of 100 bp
* 13691 16293: contig of 2603 bp in length
* 16294 16393: gap of 100 bp
* 16394 19065: contig of 2672 bp in length
* 19066 19165: gap of 100 bp
* 19166 22713: contig of 3548 bp in length
* 22714 22813: gap of 100 bp
* 22814 27165: contig of 4352 bp in length
* 27166 27265: gap of 100 bp
* 27266 32377: contig of 5112 bp in length
* 32378 32477: gap of 100 bp
* 32478 36572: contig of 4095 bp in length
* 36573 36672: gap of 100 bp
* 36673 41657: contig of 4985 bp in length
* 41658 41757: gap of 100 bp
* 41758 47024: contig of 5267 bp in length
* 47025 47124: gap of 100 bp
* 47125 52938: contig of 5814 bp in length
* 52939 53038: gap of 100 bp
* 53038 559129: contig of 6091 bp in length
* 559130 59229: gap of 100 bp
* 59230 65896: contig of 6667 bp in length
* 65897 65996: gap of 100 bp
* 65997 89427: contig of 23431 bp in length
* 89428 89527: gap of 100 bp
* 89528 95785: contig of 6258 bp in length
* 95786 95885: gap of 100 bp
* 95886 102762: contig of 6877 bp in length
* 102763 102862: gap of 100 bp
* 102863 110899: contig of 8037 bp in length
* 110900 110999: gap of 100 bp
* 110000 121002: contig of 10003 bp in length
* 121003 121102: gap of 100 bp
* 121103 133048: contig of 11946 bp in length
* 133049 133148: gap of 100 bp
* 133149 145755: contig of 12607 bp in length
* 145756 145855: gap of 100 bp
* 145856 165006: contig of 19151 bp in length
* 165007 165106: gap of 100 bp
* 165107 188243: contig of 23137 bp in length
* 188244 188343: gap of 100 bp
* 188344 194407: contig of 6064 bp in length.

FEATURES
    source
        1. 194407
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="19"
            /map="19"
            /clone="RP11-73711"
            /clone_lib="RPC1-11 Human Male BAC"
            1. 1755
                /note="assembly_fragment"
                1856..3429
                    /note="assembly_fragment"
                    /note="assembly_fragment"
                    3530..5346
                        /note="assembly_fragment"
                        5447..7338
                            /note="assembly_fragment"
                            7439..10297
                                /note="assembly_fragment"

misc_feature 10398..13590
                /note="assembly_fragment"
                13691..16293
                    /note="assembly_fragment"
                    16394..19065
                        /note="assembly_fragment"
                        19166..22713
                            /note="assembly_fragment"
                            22814..27165
                                /note="assembly_fragment"
                                27266..32377
                                    /note="assembly_fragment"
                                    32478..36572
                                        /note="assembly_fragment"
                                        36673..41657
                                            /note="assembly_fragment"
                                            41758..47024
                                                /note="assembly_fragment"
                                                47125..52938
                                                    /note="assembly_fragment"
                                                    53039..559129
                                                        /note="assembly_fragment"
                                                        559130..59230
                                                            /note="assembly_fragment"
                                                            59230..65896
                                                                /note="assembly_fragment"
                                                                65997..89427
                                                                    /note="assembly_fragment"
                                                                    89528..95785
                                                                        /note="assembly_fragment"
                                                                        95886..102762
                                                                            /note="assembly_fragment"
                                                                            102863..110899
                                                                                /note="assembly_fragment"
                                                                                111000..121002
                                                                                    /note="assembly_fragment"
                                                                                    121103..133048
                                                                                        /note="assembly_fragment"
                                                                                        133149..145755
                                                                                            /note="assembly_fragment"
                                                                                            145856..165006
                                                                                                /note="assembly_fragment"
                                                                                                165107..188243
                                                                                                    /note="assembly_fragment"
                                                                                                    188344..194407
                                                                                                        /note="assembly_fragment

BASE COUNT 50993 a 43965 c 46383 g 50458 t 2608 others
ORIGIN
Query Match 16.5%; Score 553.2; DB 54; Length 194407;
Best Local Similarity 90.4%; Pred. No. 9.6e-120;
Matches 665; Conservative 0; Mismatches 48; Indels 23; Gaps 6;
1883 gctgcctgaacatcttaagtctgtcgtgaattggagcaacaagtagagatgcca 1942
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
3297 GCTGCACGTGAACACCTTAAAGTGTGCTGTGAGTGGACCAACAAGTACAGAGATGCA 3238
1943 agaacaaggaacagcaaatgtctgtgtgtggaagtctgaacctttctggccatgaa 2002
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
3237 AGAACAGAGAAATGATCAATGCTGTGTGTGGAGAGTCTGMACTTTCGTGGCATGAA 3178
2003 ccaatataaaatcccaatatatactgaataactgaacgtctttaaatttggat 2062
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
3177 CCATTATATAAATCCCAACATATACATAAATAATATTAANAAGCTTGAANAATTGGAA 3118
2063 ttctgataccctcaatgtggccagagagacaggttggttaagatgttggtcagcaggg 2122
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
3117 TTCTGATACCTCTACTGTGGCCAGAG--ACGCTGGTAAAGATGTGGCAGACAGCAGGG 3060
2123 aagacaacaagaacaggaaggcgtgtgtgc--gcctggaactgtgctgggtttgtgtg 2181
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
3059 AACACAACAG-----GGAGAGTGTGTGTGGCCGCGCTGCGACTGTGCTGGGGTTGTTGG 3007
```





DB	1878	TCCAGTGGGCAAGACA	1894	
RESULT	10			
AF227200				
LOCUS	AF227200	3059 bp	mRNA	ROD
DEFINITION	Rattus norvegicus stauufen isoform Stau+I6 (Stau) mRNA, complete cds, alternatively spliced.			09-AUG-2000
ACCESSION	AF227200			
VERSION	AF227200.1	GI:9754865		
KEYWORDS	.			
SOURCE	Norway rat.			
ORGANISM	Rattus norvegicus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
AUTHORS	1 (bases 1 to 3059)			
TITLE	Monshausen, M., Putz, U., Rehbein, M., Schweizer, M., Desgroselliers, L., Kuhl, D., Richter, D. and Kindler, S.			
JOURNAL	Two rat brain stauufen isoforms differentially bind RNA unpublished			
REFERENCE	2 (bases 1 to 3059)			
AUTHORS	Monshausen, M. and Kindler, S.			
TITLE	Direct Submission			
JOURNAL	Submitted (20-JAN-2000) University of Hamburg, Institute for Cell Biochemistry and Clinical Neurobiology, Martinistrasse 52, Hamburg 20246, Germany			
FEATURES	Location/Qualifiers			
source	1..3059			
	/organism="Rattus norvegicus"			
	/strain="Sprague-Dawley"			
	/db_xref="taxon:10116"			
	/tissue_type="brain"			
	/dev_stage="adult"			
	1..3059			
gene	/gene="Stau"			
	227..1714			
CDS	/gene="Stau"			
	/note="RNA-binding protein; similar to Drosophila stauufen, contains four RNA-binding domains; rStau+I6"			
	/codon_start=1			
	/product="stauufen isoform Stau+I6"			
	/protein_id="AAF9B119.1"			
	/db_xref="GI:9754866"			
	/translation="MYKPVDPHSMQSYTSYSGMRGAYPPRYFYPFVPLLIVGELELVGGOFNKGKMPRPKHDAATRALRTLOSPLPERLEVNRSEENLNKSELISOVAVETALKRNLPVNEPSPILTOVARREGSPPMKKFPIVRYSGEFGVEBEGRSKITSKNAAEVALQLRLPLPLAVERKYPRIRKKSOPTEKLQAPLYGGMNTISLAIIOAKKKKEPEMLTLTERGLPRREFVMQAVVGHTTAAGAGTKKVAA RNA ENNELTIGFVKVP AQPALPKALSEKTIPVKRGDGRKRVTFEPSPGDENGNSNKDEFRMPLYSHOOLPAQ ILPWAEVAQAVGSQGHNTHKDFRAAANPAKATVTAMIARELLYGTSPTAERTILK NISSGHVHGPRTRSEDOLYYLTNRAGGVQVYKDEPKNNKKECVSLINCSSQPLVYS GIGGVDSCHDMALNILIKLSLSDDOSTEMPRIKNGVSNACRC"			
	3043..3048			
	/gene="Stau"			
BASE COUNT	817 a	778 c	749 g	713 t
ORIGIN				2 others
Query Match	37.4%	Score 1253.8;	DB 74;	Length 3059;
Best Local Similarity	86.3%;	Pred. No. 8.2e-285;		
Matches 1428;	Conservative	0;	Mismatches 207;	Indels 20; Gaps 3.
DB	454	agaagcataaccctcactgtagtaactaaatcatgcactgtgcgaactggaaaaaac	513	
DB	166	AAAAAGTAGCTTTACCATTCAGACTCATACTCACCTGTGCCTGAACCTGGAAGAANAACC	225	
OY	514	aacgtataaacctgttcgacctcacttcggaatgcagttcacctaacaacatgag	573	
DB	226	AATGATATAAGACCGCGTGACACCCCACTCTGTGATGACGATGCACCTACAGCTAATGCGATGG	285	
OY	574	aggagtgcttatcccccaggglactttaccacattccagttccacacttactatca	633	

[illegible]



```
/product="stauufen isoform Stau-16"
/protein_id="AAFP98177.1"
/db_xref="GI:9755163"
/translation="MYKVPDPSHRSMOSTYSYGMRGAVPRPYEFPVPPPLLYOVELS
VGOQFNKGKMRPPVKHDATRALRLTLOSELPRLSENGRESEENLNKSEI SOF
EALRLNLPVNEFVARRESGPPHMKNFVRVSGFVGEDEGSKISKKNRAVLEO
LRRLPLPAVERVKPRIKKSSOPTCKLOTAPYGGOMNLSLAIIOAKKREBEYM
LTERGLPRRREVMQVAVGHHTAGAGINKVAKRNAENMLLELGEKVPQAPKRP
ALKSEKTPVKKPGDGRKVTEPEPSGDESNKEDERMLYSHOOLPAQILMPV
EVAQAVGVSGHHHTDFARAAFNPAKATVTAMIAELLYGTSPTAETILKSNISGH
VPHGPRTPSEDLYYLSRAOGFOVEYKDEPKNNKNECVSLICSSQPLVSHGICKDV
ESCHDMAINILIKLSELDDQSTEMPRGNNGVSNACGR"
polyA_signal
3025..3030
/gene="Stau"
BASE COUNT      814 a      771 c      747 g      707 t      2 others
ORIGIN
Query Match      38.3%; Score 1281.8; DB 74; Length 3041;
Best Local Similarity 87.2%; Pred. No. 2.1e-291;
Matches 1428; Conservative 0; Mismatches 207; Indels 2; Gaps 2;
QY 454 aaaaacataaccctacgtgtaactaaatgacatgtaactgtgaataacttggaaaaaac 513
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 166 AGAAAGTATAGCTTCTTACCAATGAGCTCAATGACATGTGCTGAACCTGGAAGAAACC 225
QY 514 aatgtaaacctgtgtacaccttaccctgagtcagtcacacctatacaacaatgag 573
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 226 AATGTATAGCCCGTGAGCCCACTCTCGATGCACTGCACTACACTATGCGATGCG 285
QY 574 aggaagtgcttaccctcccgaggaacttaccatltccagttccacttactatca 633
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 286 TGGAGTGCTCTCCCTCCCGACAGATCTTACCCATTTCCAGTTCACCGTTACTACCA 345
QY 634 agtgaacttctgtgtgagaggaagaatttaatgcaagaagaagaagaagctgc 693
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 346 AGTTGAGCTCTCCGTGGAGAGACAGATTTAAVGGAGGAAAGATAGACACCTGT 405
QY 694 gaaacccga tgcctgtcccaagcgtctgagga tccctcgagaatgagccctcgagaag 753
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 406 GAAACTATATGCCACATGCCCGAGCCTTGAGACTCTGCAGAGTGAACCCCTTACCAAGAG 465
QY 754 gctggaagtgaaatggaagaatccgaagaagaanaa tccaataaa tctgaataagta 813
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 466 GCTAGAGGTAATGGAAGAGATCAGACAGAAACCTTAATTAATCAGAAATAAGCCA 525
QY 814 agtgtttgagattgcacttaaaaggaactgtcgttgaatttcgaagtgtgcccggagaag 873
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 526 AGTGTGTGAGATTGGCTGAAGCGGAATTGGCTGTGAATTTTGAAGTGCGCCGAGAGAG 585
QY 874 tggccacccacatgaagaacttgtgaaccaagtttcgatttggggagttgttgggga 933
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 586 TGGCCCGCCACACATGAACACTTTCTGTGACAGAGTTTCAAGTTTGGGAGTTTGTAGGGA 645
QY 934 aggtgaagggaagaagaagaatttcaagaanaaa tgcgcgcatactgttctcttgaaga 993
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 646 ACGAGAGAGGGAAGCAAGCAAGATCTCCAAGAAAGATCGCGCAAGGCGGTGTGAGACA 705
QY 994 gctgaagaagttaacggccctgtcctgtcagttgaacgaagttaagccttaagatacaaaaaga 1053
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 706 GCTAAGAGAGGTGCCACCTCTACTCTGTGTGAGCGAGCGAAGCCCAATTAAGAGAA 765
QY 1054 aacaaaacaccatagttcaagccacagaacagccagataatggccagggaatcaatccgat 1113
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 766 AAGTCAGCCCACTGCAACACTTACAGACTGCCCCGCACTATGGCCAGAGGATGATCTTAT 825
QY 1114 tagcgcaactgtgcctcagatccagcaggcaaaaaaagaagaagccagagttacacgtctct 1173
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 826 TAGCAGACTGTGACAGATCCAGACAGCAAAAAAGAGAGAGCCGAGTATACATGCTCT 885
QY 1174 ccaagaagcagagcctcccgccgcccgaaggaggtttgtatgacaggttgaagttggaacca 1233
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 886 TACAGAAACAGGTCTTCCACGTCGACAGGAGTTTGTGATGACAGGTGAAGGTGGGCATCA 945
```

```
QY 1234 cactgcagaagaagaaacggygacccaacaagaagtgcccaagcgcaatgcagcgagaacat 1293
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 946 CACCGAGAAAGAGGAGCAGGTACCAATTAAGAGGTGGCCAAACGCCAATGCTGTAGAAACAT 1005
QY 1294 gctggagatccttggtttcaaaagtcgcgagcgagcgcccaacaacccgactcaagtc 1353
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1006 GCTGGAGATCTTGGGTTCAAGTCCCTCCGCGCAGCTCGCCAAAGCATCAAAATC 1065
QY 1354 agggagaagacacccataaagaacacagagga tggagaagaagtaaccttlttgaacc 1413
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1066 AGAAGAGAAAGACTCCACTAAGAAAGAACAGAGATGGAAGGAAGTAACATTTTGTGAACC 1125
QY 1414 tggctctgggagatgaanaatlyggaactagtaataagaagatgagltcaagatgccttact 1473
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1126 TAGCCTGGGGATGAATAATGAACAGTAAAGAGATGAAGATGACAGTGCCTATCT 1185
QY 1474 aagtaatcagcaagctgcctgtcgtgaattcttcccatggtgtgcccgaagtgctgcagctgt 1533
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1186 TAGCCATCAGCAGCAGCTGCCACCCGGGATTTCTCCCATGTGTCAGAGAGTGGCCAGGCTGT 1245
QY 1534 aggaagttaatcaagagacatcacacccaagaatlttaccagggcagctccggaatctgccaa 1593
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1246 CGGGTTAGTCAGAGACACCATACCAAAAGATTTGCGCAGGGGACGCTCCAATCTTGCCAA 1305
QY 1594 ggcacagtaactgtccatgataagcccgagagtggtgtatlyggggagcactcgcccaagc 1653
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1306 GGCACAGGTAACTGCCATGATAGCCGAGAAATTTGTACGGGGGACCTCGGCCACAGC 1365
QY 1654 cgaagaccatttaagaataaacaatctctcaagcgcaagtaacccaatgagacttcaagag 1713
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1366 CGAGACCATTTTAAAGTAACATCTTTCAGGACATGTACCCCATGTGACCTTGACACAAAG 1425
QY 1714 accctctgagcaactgactatcttccagagtcacaggaatccaggttgaatacaaga 1773
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1426 ACCCTGTGACCACTGTACTACCTTTTCAGAGCCAGGATTCACAGTTTGAATACAAAGA 1485
QY 1774 ctctcccaaaaaaacaagaacgaatltgtatcttataatltgtctctcctcagacc 1833
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1486 CTTGCCCAAGAACACAGAGAGAGGTGTATCTTCATCAACTGCTTTACACAGCACAC 1545
QY 1834 tctgatacagcatgtgatgcgcaagga tgtggaagtcctctgcacatataatgctgtcgtaa 1893
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1546 TCTGTGTACAGCAATGAGTGGGCAAGAGATGTGAGTCTGTCAATGATATGCTGCACCTGAA 1605
QY 1894 catcttaaaagtgtctgtcgaagttggaaccaagaatlaagagaatccaaagaacagga 1953
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1606 CATCTTAAAGCTGCTGTCTGAGTTGACCAACAGACACAGAGATCCAAAGACAGGAAA 1665
QY 1954 cgaaccaatgtctgtgtgaggaagtgctgaaaccttctctgycacatgaacatacaaa 2013
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1666 CGGACCAAGTTTCAAGCTGGGGGAGGTGCTGAACCTATTCTGGCCAGTAACATTAATAA 1725
QY 2014 tcccaacatatatactgaaataactga -aactgcttgaanaatttgaattctgtatcc 2072
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1726 -CCCAACATATATATCCGAAATAATCTGAGAAACCTTTGAAAATTTGGAATTTCTGATAAC 1784
QY 2073 tccagatgggcgagaga 2089
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1785 TCCAGTGGGCCAAGACA 1801
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
LOCUS      AF061942      2860 bp      mRNA
DEFINITION Mus musculus RNA-binding protein stauufen mRNA, complete cds.
ACCESSION  AF061942
VERSION    AF061942.1 GI:4335944
KEYWORDS
SOURCE
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
```

```

/clone="RP3-470L14"
/clone_lib="RPCT-3"
misc_feature      3061..3077
                  /note="Weak data"
misc_feature      95583..95627
                  /note="Single clone region. pUC subcloned to give 12x
                  coverage. Assembly confirmed by restriction digest data."
BASE COUNT      33974 a 25115 c 24233 g 30514 t
ORIGIN

```

```

Query Match      40.5%; Score 1357.4; DB 34; Length 113836;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1369; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1969 gtcgtgagagtgctgaaccttctctgccaatgaacatataaaatcccaatataac
DB 34965 ggttgagagtgctgaaccttctctgccaatgaacatataaaatcccaatataac
QY 2029 tgaataatctgaactgcttgaataattggaattcttgatccctccagtgagccgagag
DB 34905 tgaataatctgaactgcttgaataattggaattcttgatccctccagtgagccgagag
QY 2089 acaacgtgtgtaagagatgtgagcaagcaaggaagacaagaacaagaagagagcag
DB 34845 acaacgtgtgtaagagatgtgagcaagcaaggaagacaagaacaagaagagagcag
QY 2149 ctgtgagcc-gagctgagactgtctgaggggttltgtlgtatgagcaactcgtgagccgt
DB 34785 ctgtgagccgagctgtgagactgtctgaggggttltgtlgtatgagcaactcgtgagccgt
QY 2208 cctctgagcaatgagcagctgtctgttggagaaagagctcccaagcagctgtgtcccg
DB 34725 cctctgagcaatgagcagctgtctgttggagaaagagctcccaagcagctgtgtcccg
QY 2268 ggaacacaaagaatcacacacccctgagcagctgtgtgtcttcttccctgtgt
DB 34665 ggaacacacaaagaatcacacacccctgagcagctgtgtgtcttcttccctgtgt
QY 2328 gaaagaagaagaagcagcagcagcccttctcaagctgtgcaactcaagacaattggagacaac
DB 34605 gaaagaagaagaagcagcagcagcccttctcaagctgtgcaactcaagacaattggagacaac
QY 2388 cctgagcagcagctgtgaggaagagccttgaacggcccaagcagctaaagaacagcagaa
DB 34545 cctgagcagcagctgtgaggaagagccttgaacggcccaagcagctaaagaacagcagaa
QY 2448 aatcaaatgcttccctactcagcgtgacccaacttctctagtgtgacagcccccacac
DB 34485 aatcaaatgcttccctactcagcgtgacccaacttctctagtgtgacagcccccacac
QY 2508 ctctctgagtaaccacacatcacacacactgtcttctctcacaagtgatctgtatctta
DB 34425 ctctctgagtaaccacacatcacacacactgtcttctctcacaagtgatctgtatctta
QY 2568 gttctatcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt
DB 34365 gttctatcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt
QY 2628 caattgtatctatgaatgaagcagcttggaaactgtctgagactgtactatataa
DB 34305 caattgtatctatgaatgaagcagcttggaaactgtctgagactgtactatataa
QY 2688 atctaacgcagaaagatcatatcatatcatatcatatcatatcatatcatatcatat
DB 34245 atctaacgcagaaagatcatatcatatcatatcatatcatatcatatcatatcatat
QY 2748 aaccgcagagcaattcaatgagatgcaaatgtgtgacctgtgtgtgtgtgtgtgtgtgt
DB 34185 aaccgcagagcaattcaatgagatgcaaatgtgtgacctgtgtgtgtgtgtgtgtgtgt
QY 2808 ggaacttcaaaaacttgaatgaatgtccacagtgatataaatttgacaatttgaat

```

```

DB 34125 GGACTTTCAAAACCTGTGAATAGTCTCCACAGTTGTATAAATGTGACAAATTTAGAAAT
QY 2868 tttaactttagatgataatattgttccatttatttatttatttatttatttatttatttattt
DB 34065 tttaactttagatgataatattgttccatttatttatttatttatttatttatttatttattt
QY 2928 aacagagcttaaatgaacttgcacatcgttctttaaagaatttaaagaattgaattgatac
DB 34005 aacagagctttaaataatgaacttgcacatcgttctttaaagaatttaaagaattgaattgatac
QY 2988 tatacatagtctcttcttgaagacttgaacttcaactacatacagaatagatccatcgtatga
DB 33945 tatacatagtctcttcttgaagacttgaacttcaactacatacagaatagatccatcgtatga
QY 3048 gtccatataaacctgcagagatgatttccagagagtgctcgtatcgttataatataatccca
DB 33885 gtccatataaacctgcagagatgatttccagagagtgctcgtatcgttataatataatccca
QY 3108 tttaggctgaaaagaatgaactgaacttctgtgtatatacagctgtgtgtgtgtgtgtgtgt
DB 33825 tttaggctgaaaagaatgaactgaacttctgtgtatatacagctgtgtgtgtgtgtgtgtgt
QY 3168 ttactgtacacagaaagtgtgtgacactgagcgtcgtgtgtgtgtgtgtgtgtgtgtgtgtgt
DB 33765 ttactgtacacagaaagtgtgtgacactgagcgtcgtgtgtgtgtgtgtgtgtgtgtgtgtgt
QY 3228 tagccctgagatgaatgaatgactgtcttccatcatttcttgaagctgtgaatttcttcccat
DB 33705 tagccctgagatgaatgaatgactgtcttccatcatttcttgaagctgtgaatttcttcccat
QY 3288 ggaatgttaagtaaaactgaatggttgcacatcaatgaatgtgaatgaatgaatgaatgaat
DB 33645 ggaatgttaagtaaaactgaatggttgcacatcaatgaatgtgaatgaatgaatgaatgaat

```

```

RESULT 8
AF290989
LOCUS      AF290989      3041 bp      mRNA      ROD      09-AUG-2000
DEFINITION Rattus norvegicus stauferin isoform Stau-16 (Stau) mRNA, complete cds, alternatively spliced.
ACCESSION AF290989
VERSION    AF290989.1 GI:9755162
KEYWORDS   Norway rat.
SOURCE     Rattus norvegicus
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 3041)
            Monshausen, M., Putz, U., Rehbein, M., Schweizer, M.,
            Desgroseillers, L., Kuhl, D., Richter, D. and Kindler, S.
            Two rat brain stauferin isoforms differentially bind RNA
            TITLE      Unpublished
            JOURNAL    2 (bases 1 to 3041)
            REFERENCE Monshausen, M. and Kindler, S.
            AUTHORS    Submitted (28-JUL-2000) Institute for Cell Biochemistry and
            TITLE      Clinical Neurobiology, University of Hamburg, University Hospital
            JOURNAL    Eppendorf, Suderfeldstr. 24, Hamburg D-22529, Germany
FEATURES
            source
            1..3041
                /organism="Rattus norvegicus"
                /strain="Sprague-Dawley"
                /db_xref="taxon:10116"
                /tissue_type="brain"
                /dev_stage="adult"
                1..3041
                /gene="Stau"
                227..1696
                /gene="Stau"
                /note="RNA-binding protein; similar to Drosophila stauferin;
                contains four RNA-binding domains; rStau-16"
                /codon_start=1

```







AUTHORS Falcon, A.M., Fortes, P., Marion, R.M., Beloso, A. and Ortin, J.  
 TITLE Interaction of Influenza virus NS1 protein and the human homologue  
 of Staufen in vivo and in vitro  
 JOURNAL Nucleic Acids Res. 27 (11), 2241-2247 (1999)  
 MEDLINE 99263025  
 REFERENCE 2 (bases 1 to 3066)  
 AUTHORS Ortin, J.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-JAN-1999) Ortin J., Molecular and Cellular Biology,  
 Centro Nacional de Biotecnologia, Cantoblanco, 28049 Madrid, SPAIN  
 FEATURES Location=Quillifers  
 source 1..3066  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 gene 36..1526  
 /gene="staufen"  
 36..1526  
 /gene="staufen"  
 /function="dsRNA binding protein"  
 /codon\_start=1  
 /product="staufen protein"  
 /protein\_id="CAB40082.1"  
 /db\_xref="GI:4572588"  
 /translation="MKGKKPMYKVPDYSHMOSTYNYNMKGAYPPRYEYFPVPL  
 LYVELSVGGQFNGKGTROAKHDAAKRLTIQNEVPERLEVNGRESEENLTK  
 SEISOVEFIALKRLPYNEFVARESPGPHKNFTKYSVGFGEQSGSKSIKKNA  
 AIAVLELKLPLPLPAYERVKPRITKKTKPIVKPQTSPEVGQINPISRLQIQAKK  
 EKEPEYTLTERGLRRREYVQYKVGHNHTEEGTKKVKARNAEMLEILQFKVP  
 QRSHTQVQGEDTHETRGKMSNLELNALQEMGTSKEDERFPYPSHOOLA  
 GILPMPEVQAVQSVQGHHTKPTRAAPNFAKTVIAMIARELLGCTSPATILK  
 NISSGHVPHPLRPSOLDYLSRQGVQYVEYDPPKNNKEFVLSINCSQPLIS  
 HGICKQVSEMDALNIIKLISLDQSTMPRTGNGPKVCGRC"  
 BASE COUNT 891 a 693 c 699 g 783 t  
 ORIGIN

Query Match 83.6%; Score 2799.4; DB 68; Length 3066;  
 Best Local Similarity 99.5%; Pred. No. 0;  
 Matches 2870; Conservative 0; Mismatches 6; Indels 7; Gaps 6;

Oy 459 gcaataccctactgtagaactaaatgactgtgcatgaacttggaaaaaaccatgt 518  
 |||||  
 Db 1 GCATAACCCCTACTAGTAATGACATGTCATGAACTTGAAAAAACCAATGT 60  
 Oy 519 ataagcctgtgacccctactctcgatgcagtcacactataactacaatgagaag 578  
 |||||  
 Db 61 ATAAACCTGTGACCTTACTCTCGATGCACTCCACTATTAACAACATGAAGAGG 120  
 Oy 579 gtgctatcccccagagactttaccatcttcagttccacacttacttacttaagtg 638  
 |||||  
 Db 121 GTGCTTATCCCGAGTACTTTTACCAATTTCCAGTTCACCTTTACTTTATCAAGTGG 180  
 Oy 639 aacttctgtggagagacaatlttaatgcaaaagaaagacaagagcgtcgaac 698  
 |||||  
 Db 181 AACTTCTGTGGAGACAGCAATTAATGCAAAAGAAAGACAAGACGCTGGCAAC 240  
 Oy 699 acgatgcgtgcgcaaaagcgttgagatccctgcgaatgagccctgcgagagaagctgg 758  
 |||||  
 Db 241 ACGATGCTGCTGCCAAAGCGTTGAGATCTGCAAGATGAGCCCTGCGCAGAGAGCTGG 300  
 Oy 759 aagtgaaatggaagaaatccgaagaagaataatcctaataatctgaataaagtcagtg 818  
 |||||  
 Db 301 AGGTAAATGGAAGAAATCCGAAGAAGAAATCTCAATTAATCTAAATTAAGTCAAGTGT 360  
 Oy 819 ttgagatgacttaaacggaactgtcgtlgaatttcgaggtggcccgagagatggcc 878  
 |||||  
 Db 361 TTGAATGCACTTAAGGAACTGCTGTAATTTGAGGTGCGCCGAGAGAGTGGCC 420  
 Oy 879 caccacacatgaagaactttgacaaagcttggttgaggaagtttgggggagaagtg 938  
 |||||  
 Db 421 CACCCACCACTGGAAGAACTTTGTGACCAAGGTTTGGTTGGGAGGGAAGGAG 480  
 Oy 939 aagggaagaagaagatttcaagaagaataatgcgcacatagctgttcttgaggagctga 998

Db 481 AAGGAAAAAGCAAGAAAGTTTCAAGAAATGCCGCATGATGCTGTGAGAGCTGA 540  
 |||||  
 Oy 999 agaagtaaccgcccctgcctcagtlgtgaagagatgaagctagaatacaaaagaacaa 1058  
 |||||  
 Db 541 AGAAGTTACCGCCCTGCTCAGTGAACGATTAACCTTAGATCAAAAGAAACAA 600  
 Oy 1059 aaccatagtaagaccacaagaacccagaatattggcccaagggaatcaatccgttagc 1118  
 |||||  
 Db 601 AACCATAGTCAGCCACAGAACACCAGATATGSCCAGGGGATCAATCGATTAGCC 660  
 Oy 1119 gactggcccatccagcagcagcaaaagaagaagagccagatgacacgtctccacag 1178  
 |||||  
 Db 661 GACTGCCCATCATCCAGACGCAAAAAGAGAGACCCAGATGACACGCTCCTCAAG 720  
 Oy 1179 agcgaagcctcccgccgcgcgaagagattgtgatgcaggtgaaggttggaaaccaactg 1238  
 |||||  
 Db 721 AGCGAGCCTCCCGCGCGCAGAGGAGTTGTGATCAGGTAAAGGTGGAACACACTG 780  
 Oy 1239 cagaagaacgagcacaagaagaagtggtgccaagcgaatgacgcagagacatgctgg 1298  
 |||||  
 Db 781 CAGAAGGAACGGGCACCAAGAAAGGTGGCCAACTGCAATGCAACCGAATGCTGG 840  
 Oy 1299 agatcctgtgttcaaagttcccgacgagcagcccaacaaccgcaactcaagtcagag 1358  
 |||||  
 Db 841 ACATCCTGTGTTTCAAAAGTCCCGCAGCG -CAGCCACCAAAACCGCACTCAAGTCAAGG 898  
 Oy 1359 agaagacaccataaagaaccaaaggtgtgaagaagaataacc-ttttgaacctggc 1417  
 |||||  
 Db 899 AGAAGACCCCTTAAGAAACACAGGGGTGGAAGAAAGTAACTTTTGTGAACCTGGC 958  
 Oy 1418 tctggagat-gaaatgtagctagtaataagaagatagatgagatgacctactatag 1476  
 |||||  
 Db 959 TCTGGGATGGAATVGGAGTACTGTAATPAAGAGATGATGATGATGATGATGATGATG 1018  
 Oy 1477 tcatcagcagctgcctgtctgtgaatcttcccatgtgtccgagctgcgcaagctgtlag 1536  
 |||||  
 Db 1019 TCATAGCAGAGCTGCTGCTGATGATCTTCCATGCTGTCGCGCAGAGCTGATAGG 1078  
 Oy 1537 agttagcaagagacatcaacaaagaatttaccagggcagctccgaatcctgtccaagc 1596  
 |||||  
 Db 1079 AGTTAGTCAAGGACATCAACCAAGATTTTACAGGGCAGCTCGAATCCGCAAGGC 1138  
 Oy 1597 caaggttaactgcatgtagagcccgagagttgttgaatggggacactgcgccagcgca 1656  
 |||||  
 Db 1139 CACGGTAAGTGCATGATAGCCCGAGAGTGTGTATGGGGGACCTGCGCCACGCCGA 1198  
 Oy 1657 gacca tttaagaataaacatctcttcagggcagctaccaccaatgacacttcaacgaagac 1716  
 |||||  
 Db 1189 GACCATTTTAAAGATTAACATCTCTTCAGGCCACGCTACCCCATGACCTTCAGCAGAC 1258  
 Oy 1717 ctctgagcaactggaactcttccagagtcgaaggttccaggttgaatacaagaact 1776  
 |||||  
 Db 1259 CTCTGAGCAACTGCACTTCTTCCAGAGTCCAGGATTCAGGTGATTAACAAAGACTT 1318  
 Oy 1777 ccccaaaacaagaagaacgaattgtatctcttaaatgactctcttcagccacct 1836  
 |||||  
 Db 1319 CCCCAAAACAAACAAGAAATTTGTATCTTATCAATTTGCTCTCTCAACCACTCT 1378  
 Oy 1837 gatcagcatgtgtatccgcaagaagatgttcgtgcaatgataatgctgtgcgtlgaat 1896  
 |||||  
 Db 1379 GATCAGCCATGATCGGCAAGAGATGTGAGTCTGATGATGATGATGATGATGATGATG 1438  
 Oy 1897 cttaaagtgtcgtctgagttgagccacaagaatcagagatgcaagaacaagaagc 1956  
 |||||  
 Db 1439 CTTAAAGTTCGCTGTGATGATGAGCAACAAAGTCAAGAGATCCAAAGAGGAAAGCG 1498  
 Oy 1957 accaatgtctgtgttggaaggtgtcgaacctttctgccaatgaacatataaacc 2016  
 |||||  
 Db 1499 ACCAATGCTGTGTGTGAGAGGTGCTGAACCTTTCTGCGCATGACATTTATAAATCC 1558  
 Oy 2017 caacatataactgaataactgaactgtcttgaanaatttgaanaattctgataactcca 2076  
 |||||



AF061940  
LOCUS AF061940 3142 bp mRNA PRI 04-MAR-1999  
DEFINITION Homo sapiens staufen protein (STAU) mRNA, alternatively spliced,  
complete cds.  
ACCESSION AF061940  
VERSION AF061940.1 GI:4335950  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Wickham,L., Duchaine,T., Luo,M., Nabl,I.R. and Desgroselliers,L.  
TITLE Mammalian staufen is a double-stranded RNA and tubulin binding  
protein which localizes to the rough endoplasmic reticulum  
Mol. Cell. Biol. 19 (3), 2220-2230 (1999)  
JOURNAL 99147057  
MEDLINE 2 (bases 1 to 3142)  
REFERENCE Desgroselliers,L., Wickham,L. and Luo,M.  
AUTHORS Direct Submission  
TITLE Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900  
JOURNAL Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada  
FEATURES  
source Location/Qualifiers  
1. 3142  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="20"  
/map="20q13.1"  
1. 3142  
/gene="STAU"  
288..1778  
/gene="STAU"  
/note="RNA binding protein"  
/note="alternatively spliced transcript"  
/codon\_start=1  
/product="staufen protein"  
/protein\_id="AADI7532.1"  
/db\_xref="GI:4335951"  
/translation="MKIGKKPMYKPVDPYSMSTVYNNMKGAVYPPRYFYPVPPL  
LYQVELSVGGQFNGKGTROAKHDAAKLRLLQNEPDERLEVNGRESEENLNK  
SEISQVEIATLRRLPVNFVARSQPPHMKNFYTKVSQVEFGEBSKSKIKNA  
AIAVLEIKLPLPAAVENVKPIRKTKPIVKPQTSPEQGINPISRLQIQAKK  
EKDEPYLLTERGLPRRREVMQVGNHTAEGTNKVKAAAEAMLEILGFKVP  
OROPTKALKSEKTPIKKPGDKRVTFEFGSGDENGTSNKKEDFPMRPYSHOOLA  
GILPMPEVNAOVNVSQGHHTKDTFRAAPNAKTVYMTARELLYGSTPAETILK  
NNISSGVHPLETPESQDLYLSRVQGFQVEYDFPKNNNEVSLINSSQPLIS  
HGIGKDVESHQDMALNLIKLSLIDQOSTEMPTGNGPMSCGRC"  
BASE COUNT 862 a 778 c 733 g 769 t  
ORIGIN  
Query Match 87.4%; Score 2926; DB 10; Length 3142;  
Best Local Similarity 93.8%; Pred. No. 0;  
Matches 3142; Conservative 0; Mismatches 0; Indels 206; Gaps 1;

Dh 241 AGCAGCC----- 247  
Oy 301 gtgactcgtcactgaactccactccactccaggtcagcagatttccactcagctc 360  
Dh 248 ----- 247  
Oy 361 ccgaatacgtgaltacagagtttataccaacttaacctcagaaactgacaagaac 420  
Dh 248 ----- 247  
Oy 421 aacattgtccctgaaagccctctttaaanaaagaacataaccctactgtagaact 480  
Dh 248 ----- -AGAAAGCATTAACCCCTACTGTGAACT 274  
Oy 481 aaatgcactgtgcatagaacttggaanaaaacatgtaagaacctgttgaccttactc 540  
Dh 275 AAATGCACCTGTGCATGAACCTTGAAANAACCAATGTATAGCCTGTTGACCTTACTC 334  
Oy 541 tcgagtcagltccacctataactacaacatgagagaggtcctatcccccaggtactt 600  
Dh 335 TCGATGCGATGCTCACCCTATTAACATGAGAGAGGTGTTTCCCGAGGTACTT 394  
Oy 601 ttaccattccagltccaccttaacttatacgaatggaacttctgtggaagacaga 660  
Dh 395 TTACCCATTTCACAGTTCCACCTTTACTTTATCAAGTGAACCTTCTGTGGAGACACA 454  
Oy 661 atttaatgcaagaagaagaagaacagcctgcgagaacacagatgtctgtccaaagcgt 720  
Dh 455 ATTTAATGGCAAGGAAGACAAAGACAGAGCTGCGAAGACGATGCTGCGCAAGCGTT 514  
Oy 721 gagatcctcgcagaaatgagccctgcagagagagcttgagtgaaatgaaatccga 780  
Dh 515 GAGATCTCTCGAGAAATGAGCCCTCCAGAGAGCTGAGAGTGAATGGAAGAAATCCGA 574  
Oy 781 agaagaanaactcctaataactgaataaagtcgaatgtgttgagatgtaactaaagaa 840  
Dh 575 AGAAGAAATCTCATTAATGAATGAATGAATGAATGTTGAGATTGACATTAAACGGAA 634  
Oy 841 ctgtcctgtgaatttcgaggtgtgtcccgagagatgtgccccacacatgaagaacttgt 900  
Dh 635 CTTCGCTGTGAATTTGAGGTGCGCCGGAGAGTGGCCACCCCAATGAAGAACTTTGT 694  
Oy 901 gaccgaagtttcgtgttgagagattgtgtgaggaaggtgaaggaagaacaagaatttc 960  
Dh 695 GACCAAGTTTGGTGTGGGAGATTGTGGGGAGAGTGAAGGAAAGACAGACAGATTTTC 754  
Oy 961 aaagaanaaagtcgcacatagtctgtcttgagagagctgaagaagttaccgcctgtcgc 1020  
Dh 755 AAAGAAATATGCCGCCCATAGCTGTTCTTGAGAGAGCTGAAGAAAGTTAACGCCCTCTGC 814  
Oy 1021 agttgaacgagtaagaagcttagaatacaaaagaacaaacataagtccaagccagac 1080  
Dh 815 AGTTAAAGAGATGAAGCCCTAGAAATCAAAAGAAACCAATGATCAACCCACACAC 874  
Oy 1081 aagccagaataatgacagagagatcaatccgaattgacgaactgagccagatccagcagc 1140  
Dh 875 AAGCCAGATATATGCGCCAGAGGAGATTAATCCGATTAGCCGACTGCGCCAGATCCACAGCC 934  
Oy 1141 aaaaaagaagaagaagcagaatlacacgctccatcaagaagcagagcctccgcgcagag 1200  
Dh 935 AAAAAAGAGAAAGAGAGCCAGAGTACAGCTCCTTCACAGAGCGAGCCCTCCGCGCAG 994  
Oy 1201 ggaatttgatgacaggtgaaggttggaagaacacactgcagaaggaacagggacacaa 1260  
Dh 995 GGAGTTTGTGATGACAGGTGAAGAGTTGAAACACACTCAGAGGAAGACGGGACCAACAA 1054  
Oy 1261 gaagttgcaacagcgaatgacgcagaaacatgtcttgagatcccttggtttcaaaagctcc 1320  
Dh 1055 GAAGTGGCCCAAGGCCCATGACAGCCGAGACATGCTGTGAAGATCCTTGGTTTCAAAAGTCC 1114  
Oy 1321 gcaagcagacccacacaaacccgcacatcaagtccagaagagaagacacccaataaagaac 1380  
Dh 1115 GCAGCGCAGCCACCAACCAACCCGACCTCAAGTCAAGAGAGAAAGACACCCATTAAGAAAC 1174

QY 1231 ccaactcgagaagaaaggggcaacaaagaagtgcgcaagcgcaatgcgagagaa 1290  
 Db 1068 CCACACTGCAGAAAGAACGGGACCAACAAGAGGTGGCCAAAGCGCAATGCACCCAGAA 1127  
 QY 1291 catgcggagatccttggtttcaaaagtcggcgagcgagcccaaaacccgactcaa 1350  
 Db 1128 CATGCGGAGATCCTTGTTTCAAAAGTCCCGGACGGCGCACGCCAACCCGCACTCAA 1187  
 QY 1351 gtcaagaggaagacacccaataaagaacaaaggaatggaagaagaatgtaacctttctga 1410  
 Db 1188 GTCAGAGGAGAACACCCCATTAAGAACCA--GGATGGAAGAAAGTAACTTTTGTGA 1246  
 QY 1411 acctggtcttggygatgaaatgggacatgaataaagaagatgagltcaagatgacctta 1470  
 Db 1247 ACCTGCTCTGGGGATGAAATAATGGACATAGTATAAAGAGATGATGATCAGATGCTTA 1306  
 QY 1471 tctaaatcaatcgagactgctgctgctgagaaatctctcccaatggtgcggaggtgcgagc 1530  
 Db 1307 TCTAACTCATCAGACACTGCTGCTGGAATCTTCCCATGTGTCGCGAGGTCCGAGGC 1366  
 QY 1531 tctagagatctcaagagacatcacacaaagaatttaccagggcgagctccgaaatctgc 1590  
 Db 1367 TGTAGAGATTAGTCAGAGACATCACACCAAGATTTTACCAGGGGACGCTCCGAATCTGC 1426  
 QY 1591 caagcgcaagtaactgcatgaatgacccgagagtgctgtglaatgggggacactgcgcaac 1650  
 Db 1427 CAAGGCACGGTAACTGGCATGATGCCGAGAGTTGTGTATGGGGGACCTCGGCCAC 1486  
 QY 1651 agccggagacattttaaagaataacatctcttcaagccagctaccccaatggaacctcaac 1710  
 Db 1487 AGCCGAGACCATTTTAAATAATACACCTTTCAGGCACTTACCCATGAGACCTCTCAC 1546  
 QY 1711 ggaagaccttgagcaactggaactatcttccagagtcgaaggaatccaggtgaatataaa 1770  
 Db 1547 GAGACCTCTGAGCACTGAGACTATTTTCCAAAGTCCAGGATTTCCAGTTTAATACAA 1606  
 QY 1771 agacttcccaaaaacaagaagaactgtatcttcaatcaatgtgctcctcagcc 1830  
 Db 1607 AGACCTCCCAAAAACAAGAACGAATTTGTATCTTATCATTAATGTGCTCTCAGCC 1666  
 QY 1831 acctctgatacagcgaatgtaatgcgcaagaatgtgaggtctctgcataatgagtgct 1890  
 Db 1667 ACCTCTGATCAGCATGATGATCGCAAGATGTGAGTGTGCTGCCAATGATGCTCGCT 1726  
 QY 1891 gaacatctaaagtgtgctgtcgtgagtgagccaagaatgataagatgtccaaagaacag 1950  
 Db 1727 GAACATCTTAAAGTGTGCTGTGAGTTGGACCAACAAGTACGAGATGCAAGAACAG 1786  
 QY 1951 aaaggaacaaatgtctgtgtgtgagaggtgtgaaccttctgtgcatgaacattata 2010  
 Db 1787 AAAGGACCAATGTGTGTGTGTGGAGGTGCTGAACCTTTCTGGCCATGAACATTATA 1846  
 QY 2011 aaatcccaacataatactgaataaatactgaactgtcttgaataatltggaattctgata 2070  
 Db 1847 AAATCCCAACATATATATGAAATACTGAACCTGTTTGAATAATTTGGAATTTGATA 1906  
 QY 2071 cctccagtgagcgcaagaagaacagtggtgaaagaatgtgagcagcgcaagcggaagaaca 2130  
 Db 1907 CCTCCAGTGGCGCGAAGACACGCGTGAAGATGTGGCGACGACGAGCAAGAACAC 1966  
 QY 2131 agaaacaagaagagcggtgtgtgagc--ggctgagactgtgtggtgtgtgtgtgtgtgtgt 2189  
 Db 1967 AGAAACACAAGAGGCGGCTGTGTGGCGGCTGAGCTGTGCTGGGCTTGTGTGTGTGTGT 2026  
 QY 2190 actcggtgacccggt 2249  
 Db 2027 ACTCGGTGACCTGTGCGGTCTCTACGCAATGACGCTGTGTGGGAAAGAGGCTGCC 2086  
 QY 2250 agccaactgtgtctcccggaagacacagcagatccacacactgtgagccctgtgtgtgt 2309  
 Db 2087 AGCAGACTGTGTCTCCGGGACACGACGATCCACACCTTGCGGACCTCTCGTGTGTGT 2146  
 QY 2310 ctttttttccctgt 2369

Db 2147 CTTTTTTTTCCCTGTGTGAAAGAAAGAACGGACGACACCCCTTCTCAAGTGTGCTACTC 2206  
 QY 2370 agacacatctggagcaaacctctgacagccatgccaagaagaagccttgcagcgcccaag 2429  
 Db 2207 AGACACTTGGGCAAAACCTTGACAGCCATGCGCAGAGAGAGGCTTTTGAACGCGCCAG 2266  
 QY 2430 agctaaagcacccagaaataatcaaatgctcttactcaagctgagcccaacttctagt 2489  
 Db 2267 AGCTAAAGACACAGAAATAATGCTTCTACTACAGCGTACCAACTTGTGTAGT 2326  
 QY 2490 gtagcagggcccaaccaactctgcaggtacccaacacatcaacatgcttctctccaa 2549  
 Db 2327 GTGCCAGGGCCGACACCTCTCGCATGACCCACACATCCACCTGCTTCTTCCAA 2386  
 QY 2550 cagtgatctgtatcttctgtatcttcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2609  
 Db 2387 CAGTGATCTGTATTTCTAGTTTCTATTTTCTTTGATGATGATGATGATGATGATGAT 2446  
 QY 2610 ttctatctgagaattctcgaatgtatctagltaaatgaagagcttggaaactgtctg 2669  
 Db 2447 TTTTCATTTGAGATTTCTCATTTGTATCTAATTAAGACAGTTTGGAACTTGTCTG 2506  
 QY 2670 agactgaacttataatlaaatacgaacgaagaatcaatcaatgtgtatgtgtgtgtgtgt 2729  
 Db 2507 AGACTGACTTTATCAATTAATCTAACCGACAAAGATCATATCCATGTGTATGTGTAGAC 2566  
 QY 2730 attttatctaatctgactaacccagagagtttcaatgagtaagtaagtggtgtgtgtgtgt 2789  
 Db 2567 ATTTTATTTTCAATTTACTAACCAGGACGTTTCAATGATGATGATGATGATGATGATGAT 2826  
 QY 2790 ttcagctgaacagctcctgtgaacttcaaaaaccttgaataagtcctccagctgtataa 2849  
 Db 2627 TTACAGCTGAAGAGTCTCTGAGCTTCAAAAACCTTAATTAAGTCTCCACAGTGTATATA 2886  
 QY 2850 attggaacattgagaattttaaacttgaatgaatgaatgtgtgtgtgtgtgtgtgtgtgtgt 2909  
 Db 2687 ATTGGAACAATTTTGAATTTTAACTTTGATGATGATGATGATGATGATGATGATGATGAT 2746  
 QY 2910 ttatatttgtatgaatgaacagagacttgaatgaatgtgtgtgtgtgtgtgtgtgtgtgtgt 2969  
 Db 2747 TTTTACTTTTGTATATCAAAACGAGCTTAATGATACCTTGTATCTGTGTATTAAGATTAT 2806  
 QY 2970 taaaaaacattgtgtatctatataatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3029  
 Db 2807 TAAAAAACATTTGTGTAT 2866  
 QY 3030 ggaatgatactcaatgtatgtatcaataaactgtcagagtgatgttccagagtgctgata 3089  
 Db 2867 GATATGATCTCCATGTAGTCCATATAAACCTGCAAGATGATTTTCCAGAGTGTCTGATA 2926  
 QY 3090 ctgttaatacatctcgaatgaagctgaagaagaatgaacatgcttctgtataacagctgt 3149  
 Db 2927 GTGTATTAATCATCTCCATTAAGGCTGAAGAGATACCTTCTTGTATACACTGT 2986  
 QY 3150 gtctccttgaatgtgtgtatctgtatcaacagaagtgtgtgtgtgtgtgtgtgtgtgtgtgt 3209  
 Db 2987 GTTGCTTTTGTATGT 3046  
 QY 3210 tccgtatgaaaaactgt 3269  
 Db 3047 TCCGTATGGAAGACCTGT 3106  
 QY 3270 tggaaatttctcccaatggaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 3329  
 Db 3107 TGGAAATTTTCTCCCATGGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 3166  
 QY 3330 aatacaataaaaaaaataa 3348  
 Db 3167 AATACTTAAAAAATAAAAA 3185

OY	3263	tttaacgcgggaatttcttcoccatgagatgtgaagtaaactiaagtgttgatcattat	3322
Dd	3421	TTTAGCGTGGGAATTTTCTTCGCCCATGCAATGTAAATAACTTAAGTGTTTGCATCAAT	3480
OY	3323	aaatgtaataactcaaaaaaaaaaaaaa	3348
Dd	3481	AAATGCTAATACATAAAAAAAAAAAAAA	3506
RESULT	4		
AK025519			
LOCUS			
DEFINITION		AK025519	PRI 29-SEP-2000
ACCESSION		AK025519	Homo sapiens cDNA: FLJ21866 fts, clone HEP02379, highly similar to AF061938 Homo sapiens staufen protein (STAU) mRNA.
VERSION		AK025519.1	GI:10438059
SOURCE			oligo capping; fts (full insert sequence).
ORGANISM			Homo sapiens hepatoma cell_line:HepG2 cDNA to mRNA, clone_lib:HEP clone:HEP02379.
REFERENCE			Homo sapiens
AUTHORS			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE			1 (sites)
JOURNAL REFERENCE			Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S. NEBO human cDNA sequencing project Unpublished (2000)
AUTHORS			2 (bases 1 to 3194)
JOURNAL REFERENCE			Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y. Direct Submission Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sunio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-Ku, Tokyo 108-8639, Japan (E-mail:cdna@elms.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT			NEBO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert Sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing; Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).
FEATURES			Location/Qualifiers
SOURCE			1..3194
			/organism="Homo sapiens"
			/db_xref="taxon:9606"
			/cell_line="HepG2"
			/cell_type="hepatoma"
			/clone="HEP02379"
			/clone_lib="HEP"
			/note="Cloning vector pME1BSFL3"
			1..3194
			/note="Highly similar to AF061938 Homo sapiens staufen protein (STAU) mRNA"
BASE COUNT			894 a 787 c 732 g 781 t
ORIGIN			
Query Match			88.4%; Score 2961; DB 34; Length 3194;
Best Local Similarity			95.4%; Pred. No. 0;
Matches 3167; Conservative			0; Mismatches 15; Indels 137; Gaps 6
OY	33	cttcagcglttgcg--ggcggctgcgcgtctctcgcgctccgcgttccttgagccg	90
Dd	1	CTTAGCGATTGCGCGCGGGGCTGCGGCTCTCTCGCGCTCCGCTTGTAAOCGCC	60
OY	91	tcccccccgc	150
Dd	61	T-CGCCCGCCGGCGCGCGCGCGCTCCTCCACAGGCGCACCTCGCGCTCTTCCCT	119

[illegible]

|||||  
Db 1201 ATCAAAAAAACAACAAACCCATAGTCAAGCCACAGACAGCCCAAGATATGCGCCAGGAG 1260  
Oy 1103 atcaatccgattagccgacttggcccgatccagcaaggcaaaaagagaagagccagag 1162  
Db 1261 ATCAATCCGATTAGCCGACTGGCCCGATGCCAGACAGCAAAAAAGAGAGAGCCAGAG 1320  
Oy 1163 taacgctcccaagagcagagcctcccgccgagagagtttga tgcagtgaaag 1222  
Db 1321 TACAGCCTCCTCAACAGAGGAGGCTCCCGCGCCGAGGAGTTGTGATGACGTGAAG 1380  
Oy 1223 gtctgaataaccaacttgcagaagaagggcaccacaagaaggttgccaaagcgaatgca 1282  
Db 1381 GTTGGAAACCACTGCACAAGAGAAAGGGGACCACCAAGAGGTGGCAAGCCCAATGCA 1440  
Oy 1283 gccgaagaatcgtcggagatcccttgg ttccaaagtcgcgagcgagccacaaccc 1342  
Db 1441 GCGGAGACATGCTGGAGATCTCTTGTTCAAAAGTCCCGACGCGACGCCCAACCC 1500  
Oy 1343 gcactcaagtcagagagagaacacccataaagaacacagggatggaagaaagttaac 1402  
Db 1501 GCACCTCACTCAGAGAGAGAAACACCATTAAAGAAACAGGGGATGGAAGAAAGTAAAC 1560  
Oy 1403 tttcttgaacctggctc tgggga tgaanaa tgggactagttaataaagaga tgaagttcaag 1462  
Db 1561 TTTTGTGAACCTGGCTCTGGGGATGAAATGGGACTAGTAAATAAGAGATGAGTTCAGG 1620  
Oy 1463 atgacctatcagta tcaagagagccttgc tggaa tcttccca tgg tgcggagtc 1522  
Db 1621 ATGCTTATCTAAGTATATAGCAGCTGCTGCGAATCTTCCATGGTGGCCGAGGTC 1680  
Oy 1523 gccagagctgtagaggttagtcaagagatacacaaccaagatttcaacggcgagctccg 1582  
Db 1681 GCCAGGCTGTAGGAGTTAGTCAAGACATCAACCAAGATTTTTCAGGGGAGCTCCG 1740  
Oy 1583 aatcttgcgaagccaggttaactgcacatgatagcccgagagttgttga tggggagcc 1642  
Db 1741 AATCCGCCCAAGGCCACGGTAACTGCCATGATAGCCGAGAGTTGTTATGGGGCAGCC 1800  
Oy 1643 tggcccaagccgagagacatttaagaa taaatctcttcaggccacgttacccca tga 1702  
Db 1801 TCGCCACAGCCGAGACATTTTAAAGAAATACATCTTTCAGGCCACGTACCCCAATGGA 1860  
Oy 1703 cctctcaagagacccctc tgaagcaactgc tcttccagag tccagagatccagct 1762  
Db 1861 CCTCTACGAGACCTCTAGACACTGACTATCTTCCAGAGTCCAGGAGATTCCAGTT 1920  
Oy 1763 gaatacaagacttcccaaaaaacaagaacgaa tttgtatctctta tcaattgtccc 1822  
Db 1921 GAATACAAAGACTTCCCAAAAAACAAGAACGAATTTGTATCTGTATCAATTGCTCC 1980  
Oy 1823 tctcagccacccctc tga tgcata tgcgcaagagatg tggagttccgcca tgaatg 1882  
Db 1981 TCTGAGCCACCTCTGTATCGCATGTATCGCCAAAGATGTGAGTCCATGATATG 2040  
Oy 1883 gctgcgctgaacaacttaagttgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc 1942  
Db 2041 GCTGGCCTAAACATCTTAAAGTTGCTGTGACTTGACACCAACAAAGTACAGAGATGCCA 2100  
Oy 1943 agaacaagaagaacaa tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc 2002  
Db 2101 ACAACAGGAACGACCAATGTCTGTGTGGAGGTGCTGAACCTTTTCTGGCCATGAA 2160  
Oy 2003 ccaattataaattcccaacata tactga aatactgaaactgc tttgaaat tlgaaat 2062  
Db 2161 CCATTATATAAATCCAAACTATATATCTGAATAATACGAAACCTGTTGAATAATTTGGAAT 2220  
Oy 2063 tctctataccctcag tggcgagagagacacggt tgg taaagagat tggggcacacagag 2122  
Db 2221 TTCTGATACCTCAGTGGGCCAGAGACACGGTGGGTAAAGATGTGGGCGACGAGAGG 2280  
Oy 2123 aagacaagaagaacaaagagcggt tgc tgc cggc tgc agctgc tgc tgc tgc tgc 2182  
|||||

Db 2281 AAGACAAGAAACACAGAGAGCGGCTGTGGCGGCTGGACTGTGGGCTTTGTTGT 2340  
Oy 2183 gatggcactcgt tgaacctggcggt tccctcaacatagcagctgc tgt tgggaaagaag 2242  
Db 2341 GATGGCCTCGGTGACCTGGGGCTCCCTAGACGAATAGCAGCTGCTGTGGGCAAGAAG 2400  
Oy 2243 gctgccagcagctgc tctcccggaacaacagagatccacaaccttgggaacctcgt 2302  
Db 2401 GGTGCCACACAGCTGTTCTCCCGGACACCGAGATCCACACCTCGGGCACCTCCGT 2460  
Oy 2303 gtctgtcttttttccctctgt tgaagaagaacggcagacaccccttcaagctg 2362  
Db 2461 GTTGTGCTTTTTCCTCTGTGAAAGAAAGAGAGGACGACCTCTCTCAACCTGG 2520  
Oy 2363 ctcaactcaacacat tgggaacaaaccttgcagacatgc cagaagagagcctttagcg 2422  
Db 2521 CTACCTCAGACATTTGGGACAAACCTTGACAGCCATGCAAGAGAGGCCCTTTACCG 2580  
Oy 2423 gccccagctaaagcacccagaga aaatcaaatgcttcc tactcagctgc tgcacaa 2482  
Db 2581 GCCCCAGACTTAAAGCACACAGAAATCAAAATGCTTCTACTCAGCGTGACCCAACTT 2640  
Oy 2483 tcttag tgcgcagcgcccaacacactcctgcagtagtcccaacacacacacactgcttct 2542  
Db 2641 TTTTACTGTGCCAGGCCCCACACCTCCTGCAGTACCCACACATCACACTGCTTTCT 2700  
Oy 2543 ctccaacagtagatcgtat tcttagttcaatatttctt tgc tgc tgc tgc tgc tgc 2602  
Db 2701 CTTCACACAGTATCTGTATCTTACTTTCATTTATTTTCTTTGATGTGATACACTAT 2760  
Oy 2603 ataaaa tttcatt tgaagatttcaaat tgc tgc tgc tgc tgc tgc tgc tgc tgc 2662  
Db 2761 ATAAATTTTCAATTTAGAAATTTCTCAATTTGATAGTTAAATATACAGATTTGGAAC 2820  
Oy 2663 tgc tgc agacgcactt tcaataa tcaacccaacgaagaatca tcaatgc tgc tgc 2722  
Db 2821 TTTGTGAGACACTGACTTTTATCATATATCAATATCAACCGAAAGATCATCTCATGTGTGTG 2880  
Oy 2723 gttagacatttatttca ttaactga ctaaccagagagtttca g tga tgc aaat tgc tgc 2782  
Db 2881 GTTAGACATTTTATTTTCAATTGACTAACCAGAGACGTTCACTGATGCAAAATGTGTGC 2940  
Oy 2783 cctctgttcagctga aagaag tctc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc 2842  
Db 2941 CCTGTGTTACAGTGAACAGTCTGACTTTCAAAACCTTGAATAAGTCTCCACAG 3000  
Oy 2843 tgtataaattggaacatttaggaatttcaact tgc tgc tgc tgc tgc tgc tgc tgc 2902  
Db 3001 TGTATTAATTTGACATTTTAGCAATTTTAAACTTTAGATGATCATTTGGTCCATTTTAA 3060  
Oy 2903 tttcaatttattt tgc ttaatgc aacagagacttaatga aacttgc tgc tgc tgc tgc 2962  
Db 3061 TTTCAATTTTATTTTGTAAATGCAACAGGACTTAAATGAACCTTGATCTCGTTTAA 3120  
Oy 2963 agattataaaaaacattgt tca taca taca tgc tgc tgc tgc tgc tgc tgc tgc 3022  
Db 3121 AGATTATTAATAAACATTGTGATCTATATCATATGTGCTCTTGAGGACTTAAGCTTCAC 3180  
Oy 3023 cactacagatatgat tccca tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc 3082  
Db 3181 CACTACAGATATGATCTCATGTACTCATATAAACCTGCAGAGAGTATTTTCCACAGTG 3240  
Oy 3083 ctgcagactgttaattacat tcccat tgaagcctgaa aaaaatgc aacttgc tgc tgc 3142  
Db 3241 CTGAGATATGTTAATATACATCTCCATTTAGGGCTGAAAAGAAATGACCTACGTTTCTATA 3300  
Oy 3143 cagctgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc 3202  
Db 3301 CAGCTGTGTTGCTTTGATGTGTGTGTACTGTAACAGAAAGTGTGCACTGAGGCTTG 3360  
Oy 3203 cgtgtg tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc 3262  
Db 3361 CGTGTGTCCGATGGAACCGTAGCCCTTGCGAGTTAAGTACTGCTTCATTCATTTG 3420  
|||||

[illegible][illegible]



1081 aagcccaaatatgagcgaaggatcaaccgaatgaagccagctggcccgagatccagcaggc 1140  
1141 aaaaaaggaaggaagcgaagatgacagctctctcaagagcgagagctcccgccgag 1200  
1010 AAAAAAGGAAGGAGAGCCAGATACAGCTCTCCACAGAGCGAGGCTCCCGCGCGAG 1069  
1201 ggaattctgatagcaagtgaagctggaaccacactgcaagaaggaagcagcagacaa 1260  
1070 GGAGTTTGATGATGCAAGGTGGAAGTGGAAACCACTGCAAGAGAGACGGCCACACAA 1129  
1261 gaagtgagcgaagcgcaatgagcagcgaagacabtgctggaatcttggttccaagtccc 1320  
1130 GAAGTGCGCCAAAGCCCAATGACGCGAGAAACATGCTGAGATCTTGTTTCAAGTCCC 1189  
1321 gaagcgagcccccacaaaccgagctcaagttagagaggaagacccaataagaacc 1380  
1190 GGAGCGGCGAGCCACCAACCCGCACTCAAGTCAAGAGAGAGAACACCCATAAAGAACCC 1249  
1381 aagggatggaagaaagtaaccttttgaacctggctctggggatggaagaatggagctag 1440  
1250 AGGGATGGAAGAAAGTAACCTTTTAACTTGGCTCTGGGGATGAAATGGACTAG 1309  
1441 taataaagagaagatgataagatgacctatactaaagtcacagcagctgctgctggaat 1500  
1310 TATAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1369  
1501 tctcccaatgagtgccggaagtgccggaagctgttagagttagtcaagacatcaaccaa 1560  
1370 TCTTCCCAATGAGTCCGAGAGTGGCCGAGCTGTAGAGTGTAGTGTAGTGTAGTGTAG 1429  
1561 agattttacaaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1620  
1430 AGATTTTACAGAGGCGAGCTCCGATCTCTGCAAGGCCAGGTAAGTGTATGATGATGATG 1489  
1621 agagtgtgtatgagtgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1680  
1490 ACAGTGTGTATGAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 1549  
1681 ttcaagcgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1740  
1550 TTTACAGGCCAGTACCCGATGACCTCTGACGAGACCTCTGACCAACTGTGATGATGATG 1609  
1741 cagagtcgaaggaatccaggttgaatatacaagaactcccaaaaaaagaagaagaat 1800  
1610 CAGAGTCGAGGATTCAGAGTTGAATACAAAGACTTCCCAAAAAACAAAGACGAATTT 1669  
1801 tgalactctatcaatgtgctctcagcagcagcagcagcagcagcagcagcagcagcagcag 1860  
1670 TGTATCTTTATCAATTTGCTCTGCTGACGACCTGTGATCAGGCAATGATGCGCAAGGA 1729  
1861 tgtggaagtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1920  
1730 TGTGAGTCTCTCCATGATGATGCTGCTGGAACATTTAAAGTGTCTGCTGATGATGGA 1789  
1921 ccaaaagaatagaagaatagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1980  
1790 CCAACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1849  
1981 ctgaagcctttctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2040  
1850 CTGAACCTTTCTGCGCATGAAACATTTAAATCCCAACATTTATTTATCTGAAAAATCTGA 1909  
2041 aactgcttgaaaaatttgaattctgatactcagtgagtgagcaggaagacagtgagta 2100  
1910 AACTGCTTTGAAAAATTTGAAATTTTGATACCTCCAGTGGCGAGAGACAGCGTGGGTA 1969  
2101 aaggaatgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2160  
1970 AAGGATGTGGGAGCAGCAGCGAGAGACCAACAGAAACAGAGGCGGCTGTGGCGGCT 2029  
2161 ggaactgtgctgagggcttctgtgtgatagagcagcagcagcagcagcagcagcagcagcag 2220

2030 GCAGTGTGCTGGGGTGTGTGATGATGAGCCACTGCGATGAGCTGGCGGTCCCTACGCAATAG 2089  
2221 cagctgctgctgaggaagaagaagcctgcccagcagctgctgtctccggagacacagcaga 2280  
2090 CAGCTGCTGTGGGAGAGAGGCTGCCAGCGAGCTGGTCTCCGGGACACACAGAGA 2149  
2281 tccacaacctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2340  
2150 TCCACACCCCTGGGACACTCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2209  
2341 gaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2400  
2210 GCAGAGACCCCTTCTCAAGCTGAGCTGCTCAGACATTTGGAGAACCCCTGGACACCAT 2269  
2401 gccaagagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2460  
2270 GCCAGAGAGAGGCTTTGAGCCGCGCCAGAGCTTAAAGCCAGAGAGAGAGAGAGAGAGAG 2329  
2461 cctactcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2520  
2330 CCACTACAGCGTACCCCACTTTTCTAGTGTGCGACAGCGCCACACCTCTCTGATAC 2389  
2521 caac 2580  
2390 CACACATCACACACTGCTTCTCTTCCAAACAGTATCTGATTTGATTTGATTTGATTTGAT 2449  
2581 ctttgatgatagac 2640  
2450 CTTTGTATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 2509  
2641 ttaataagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2700  
2510 TTTAATAGCAGATTTGGAACCTTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2659  
2701 agatcatatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2760  
2570 AGATCATATCAGTGTATGT 2629  
2761 ttcaagatgagcaaatgtgtgctgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2820  
2630 TTTAGT 2689  
2821 ccttgataagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2880  
2690 CCTTGAATAGT 2749  
2881 tgalacttgatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2940  
2750 TGTATCATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2809  
2941 tgaacttgatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3000  
2810 TGAACCTTGTATCTGT 2869  
3001 ctggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3060  
2870 CTTGAGGACTTACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 2929  
3061 tgaagatgatacttccagagtgctcgatagctgtaataatcaatccatccatccatccatccat 3120  
2930 TGCAGAGTATTTTCCAGAGT 2989  
3121 gaatgaactagcttctgatacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3180  
2990 GAATGACCTTACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 3049  
3181 aagtgtgtgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3240  
3050 AAGTGTGTGACAGCAGCTGT 3109  
3241 taagtactgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3300



Query Match	Best Local Similarity	96.1%	Score 3076;	DB 10;	Length 3217;
Matches 3217;	Conservative 1;	0;	Mismatches 0;	Indels 131;	Gaps 1;
QY	3120	agaatgaacctaaagttcttcgtatatacagcgtgttcttcttgatgtgtgttactctaacca	3179		
Db	3121	AGAAATGACCTAGCTTTCGTATACAGCGTGTGTGCTTTTGATGTTGTACTGTACACA	3180		
QY	3180	gaagtgctgcactaaagctctgcgtgtgtgtccgatgaacaaaccttgtaacccctgcag	3239		
Db	3181	GAAAGTGTGTGCACTAGGCTCTGCGTGTGTGTCCGATGTGAAACCTGTAGCCCTGCGAG	3240		
QY	3240	ttaagtaactgcttcattcatatgtttatcacgcgtgaaatttctcccatggaatgtaagta	3299		
Db	3241	TTAACTACTGCTTCATTCATGTTTACGTCGTGAATTTTCTCCCATGGAATGTAACTA	3300		
QY	3300	aaactaagtgctgttcatacatataatgttatactataaataaataaataa	3348		
Db	3301	AAACTTAAGTGTGTTCATCATATAAATGTATACTAAAAAATAAATAAATAA	3349		
RESUT.. 2					
AF061938					
LOCUS					
DEFINITION		AF061938	3217 bp	mRNA	PRI 04-MAR-1999
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
FEATURES					
source					
gene					
CDS					
BASE COUNT					
ORIGIN					

QY	1	acctctcgccgagctgtgvggcgcctctgagcgctcttaagtggtttgcgvcgagctgcgagt	60
Db	1	ACTTCTCCCGCGGCTGGGGGCGCTGAGCGCTTTAAAGGCTTTGGCGGGCGGCTTCGGCT	60
QY	61	ctctctcgagctccgcgttctcttgaacgcgtctcccccgcgcgcgcgcgcgcgcgc	120
Db	61	CTCTCTCGGCTCCGGCTTCTCTTTGACCGCTCCGCCGCCCGGGCGGGCGCGCGCTC	120
QY	121	ctcaacgagcaactcgcgcctctctccctctctgtctccctctctctctctctcttctct	180
Db	121	CTCCACGGCGCACTCCGCGCTTCCCTCCCTCCCTCCCTCCCTCTCTCTCTCTCTCTCT	180
QY	181	ctctctccctctctctgcgcgcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	240
Db	181	CTTCTCTTCCCTCTCTCCCGCCACCGCCACGAGCGCCCGCGGGGAGGATCCGGAGC	240
QY	241	agcagccagcagcagcagctgtagtttctctctgtcgccagcgtgtagtgcagtgagc	300
Db	241	AGCAAGC-----	246
QY	301	gtgactctgcgtctcaactgcacctccacctccaggtcagcgatttccactcagctctc	360
Db	301	gtgactctgcgtctcaactgcacctccacctccaggtcagcgatttccactcagctctc	360
QY	361	cgcgaataagctgtagtctcagagttatctaaaccttaacctctcagaacttgaacaagac	420
Db	361	CGGTAAGCTGAGTCTCAGAGTTATCTAAACCTTAACCTCTCAGAACTGACAAAGAC	420
QY	421	aacatgtctctgtagcgcgcctcttlttaaaaaagaagacataaccctactgtagaact	480
Db	421	AACATGTCTCTGAGCGCCCTCTTTTAAAAAAGAAAGCATACCCCTACGTGTAGACT	480
QY	481	aaatgcactgtgcctggaactctggaaaaaaaccaatgataagcctgttgacccttactc	540
Db	481	AAATGCACCTGTGCATGTAAGCTTGAATAAAACCAATGTATTAACCTGTGTGACCTTACTC	540
QY	541	tcgagtgtagtcacacctaactaactaacaacatgagagagagtgcttactccccgcgagtgact	600
Db	541	TCGAGTGCAGTCCACTATTAACACTGACATGAGAGAGGAGTGTATCCCCCGAGGTACTT	600
QY	601	ttaccacttccagttccacacttacttcttaacatgtagaactctctgtgvggagcagca	660
Db	601	TTACCACATTTCCACTTCCACCTTTACTTTTTCAGGTGGAACCTTTCTGTGGGAGCACCA	660
QY	661	atttaatgcagaagaagaacagacagcgtctgcgaacacgactgcctgcgcaaacgct	720
Db	661	ATTTATATGGCAAGAGAAAGACAACACAGCGCTGCGAACAACAGATCGTGTGCCAAGCGTT	720
QY	721	gagagctctgtagaactgagccctgcgcagagagctgtagagtgaatlgagaagaaatccga	780
Db	721	GAGGATCTCTGCAGATGAGGCCCTCGCAGAGAGCGTGGAGTGTAATGGAAGAAATTCGGA	780
QY	781	agaagaanaatctcaataaactctgaataaactgaatgaatgtagatctgaacttaaacggaa	840
Db	781	AGAAATAAATCTCAATAAATCTGAATAAATCTCAAGTGTTTGAGATTGCACTTAAACGGAA	840
QY	841	cttgctctgtagaattcgaagtgtagccgcgtagagtgtagccacccacatagaagaacttgt	900
Db	841	CTTGCTCTGTGAATTTGCAAGGTGGCCCGGAGAGATGGCCACCCACATGTAAMAACCTTGT	900
QY	901	gaccgaagtttcggtttgvggagtttctgvgvggagagtgtagggagaagaagaagaatttc	960
Db	901	GACCAAGGTTTCGGTTGGGGGAGTTTGTGGGGGAGGTGAAGGGGAAAAGCAAAATAATTTTC	960
QY	961	aaagaanaatcgcgcatacgtcttctctgtagagctcgaagaagtltaccccccctgcgcg	1020
Db	961	AAAGAAATAATCCGCCCAATAGCTGTTCTTGAAGAGCTGGAAGAAGTTATCCGCCCTCTCGTC	1020
QY	1021	agctgaaacgagtaaagcctcagaatcaaaaaagaacaaacccatagtctaaagcagac	1080
Db	1021	AGTTGAAAGCACTAAAGCCTGAAATCAAAAAAGAAACCAAAACCCATAGTCAACCCACAGAC	1080

|||||  
Db 901 TCACCAAGGTTTCGGTTGGGGAGTTGTGGGGAGAGTGAAGGAAACCAAGAAATTT 960  
QY 960 caaagaataatccgcataagctgttcttgagagcttgagaagttaacgcccctgcg 1019  
Db 961 CAAGAAAAATCCCGCATAGCTGTCTTGAGAGCTGAAGAAATTACGCCCTGCTCG 1020  
QY 1020 cagttgaacgaatlaaagcctagaatcaaaaagaacaacccatagtcgaagccaga 1079  
Db 1021 CAGTTGAACGAGTAAAGCTTAGATCAAAAAAGAAACCAACCCATAGTCAAGCCACAA 1080  
QY 1080 caagcccaagataatggcaggagatcaatccgaatlagccgaatgcccgaatccagcag 1139  
Db 1081 CAAGCCCAAAATATGGCCAGGGGATCAATCCGATTAGCCGATGGCCAGATCCACAG 1140  
QY 1140 caaaaaagagaagagcagaagttacagctctcacaagaagagagccctccggcgca 1199  
Db 1141 CAAAAAGAGAGAGAGGAGCCAGAGTACAGCTCTCACAGAGGAGGCTCCGCGCGCA 1200  
QY 1200 gggagtttgatgacagtggaagttggaacacacatgcagaaggaagcggaaccaaca 1259  
Db 1201 GGGAGTTTGATGACAGTGAAGTTGGAACCACTGCAGAGAGAGAGGGCGACCAACA 1260  
QY 1260 agaagtgagcgaagcgaatgcagccgagaacatgcttgagatccttggttcaaaagtc 1319  
Db 1261 ACAGAGTGCCCAAGGCCAATGCAGCCGAGAACATGCTGGAGATCTTGGTTCAAAAGTCC 1320  
QY 1320 gcagcagagagcccaaccccgacatcaagtcaagagaagaaacccataaagaaac 1379  
Db 1321 CCACAGGGAGGCCCAACCCGCACTCAAGTCAGAGAGAGAACCCATTAATAAAGAAC 1380  
QY 1380 caggagatgaagaagaaagtaaccttctgaacctgctcctgaggaatgaanaatgagacta 1439  
Db 1381 CAGGGAGTGAAGAAAGTAACCTTTTGAACCTGCTCTGGGATGTAATAATGGGACATA 1440  
QY 1440 gtaataaagaagaatgaatcagaatgacctatcctaagtaacgaagcctgctcgtgaa 1499  
Db 1441 GTAAATAAAGAGATGAGTTGAGAGATGCCCTTAATTAAGTATGACGAGCTGCTGAGAA 1500  
QY 1500 tctctcccaatgagcccgagctgcgagcgctgtagaagtgtgaagaacatcaacca 1559  
Db 1501 TTCTTCCCATGGTGGCCGAGGTGCCAGGCTGTAGAGATTAGTCAAGAACATCAACACA 1560  
QY 1560 aagatattcaacagagcagctccgaatccctgcgaagccaggaacaaatlttaagaataacatc 1619  
Db 1561 AAGATTTTACAGGGAGCTCGAATCTCGCAAGGCCAGGTAATGCTCATGATATGCC 1620  
QY 1620 gagaatctgtatgaggggacactgcgccacagccgagacaaatlttaagaataacatc 1679  
Db 1621 GAGAGTTGTTGATGGGGCACCTGCCCAACAGCCAGACCAATTTTAAGAAATAACATCT 1680  
QY 1680 ctctcagagcagctaccccaatgaacctctcagaagacctctgaagcaacgagtaacctt 1739  
Db 1681 CTTTCAGGCCACGTACCCCTGAGACCTCTCAGAGACCTCTGAGCAACCTGAGTATCTTT 1740  
QY 1740 ccagagtcagaagatccagcttgaaatacaagacttccccaaaaacaacaagaagaaat 1799  
Db 1741 CCAGAGCTCCAGGAGTTCCAGGTTGAATACAAAGACTTCCCAAAAAACACAGAGAGAT 1800  
QY 1800 ttgtatctctatcaaatgtctccctcagccacacttgatcagcaatggtatctgcgaaag 1859  
Db 1801 TTGTATCTCTATCAATTGCTCTCAGCCACCTCTGATGATGACCATGATGCGCAAG 1860  
QY 1860 atgtgagatcccgccatgataatgctgcgtgaacatcttaagtctgctgagttg 1919  
Db 1861 ATGTGAGTCCGTCATGATGATGCTGCGCTGACATCTTAAAGTGTGCTGATGATG 1920  
QY 1920 accaacaagaatcagagatgccaagaagaagaagaccaaatgtctgtgttgagaggt 1979  
Db 1921 ACCAACAAAGTACAGAGATGCCAAGACAGGAACGAGCAATGCTGTGTGGAGGT 1980  
QY 1980 gctgaaccttcttgccatgaacatlaaatcccacaataactgaanaatctg 2039  
|||||

Db 1981 GCTGAACCTTTCTGCGCATGAAACATTATAAATCCAACATATATACTGAATAACTG 2040  
QY 2040 aaactgcttgaanaatttgaattcttgatctcagctgagggccgagagacagtgagtt 2099  
Db 2041 AAATGCTTTTGAATAATTTGGAATTTTCTGATACCTCCAGTGGCCGAGACACAGGTGG 2100  
QY 2100 aaaggaatgagcagcagcaggaagaacaagaacaagaagagcgctgtgcgagc 2159  
Db 2101 AAAGATGTGGGACAGCAGGAGAGACAAAGAAACACAGAGAGCGGCTGTGGCCGCG 2160  
QY 2160 tggactgtctggggtttgttgatgagccactcggatgaccttgaggggtccctacgaata 2219  
Db 2161 TGGACGTCTGGGCTTTTGTGATGGCCACTCGGTGACCTGGCGCTCTTACGCAATA 2220  
QY 2220 gcaagctgctgtgaggaagaagagcgctccagcgaactgcttctccggaacccaagcag 2279  
Db 2221 GCAGCTGCTGTGGGAAAGAGGCTGCCAGCCAGCTGTTCTCCGGGACACCGCAG 2280  
QY 2280 atccacaacctgagcactcgtgttggtcttcttccctgtgtgaagaagaac 2339  
Db 2281 ATCCACACCCCTGGGACCTCCGTTTGGTCTTTTCCCTGTGTAAGAAAGAAAC 2340  
QY 2340 ggaagacccctctcaagctgtcactcagaacacattgggaacaacccctgagagcca 2399  
Db 2341 GGCACACCCCTTCTCAAGCTGGCTCACACACATTTGGGCAAAACCTGGACAGCA 2400  
QY 2400 tgcagaagaagccttgaacggcccgagactaaagaacccaagaagaatacaaatgct 2459  
Db 2401 TGCAGAGAGAGGCTTTTACGGGCCCCAGAGCTTAAGACACCCAGAAATCAATGCT 2460  
QY 2460 tctactcagcgtgacccaacttctctagtggtgcagcgccccaacacactctcagtag 2519  
Db 2461 TCTACTCAGCGTAGACCAACTTTTCTAGTGTGCCAGGCCACCACTCGCGAGTAC 2520  
QY 2520 ccaacacatcaacacactgcttctctcacaacagtgatcctgattcttcaatctatct 2579  
Db 2521 CCACACATCACACACGCTTCTCTCAAGTGAATGCTGTCTTCAATTTCAATATTT 2580  
QY 2580 tcttttgatgataagacataaanaatttcaattgaagaattctcaatgtatcta 2639  
Db 2581 TCTTTGATTGATAGACCTATATTAATTTTCAATTTAGAAATTTCTCAATTTGTATCTA 2640  
QY 2640 gttaaaatgaacagtttgaaacttgcagagacttcaatlaaataatcaaacagca 2699  
Db 2641 GTTAATATACACAGTTTGAACACTGTGTGAGCTGATTTATCAATATCAACGACA 2700  
QY 2700 aagatcataatcaatgtatgtgtgtagaacaatttcaatctgaactgaacagagcag 2759  
Db 2701 AAGATCATATCCATGTGTGTGTGAGACATTTTATTTCAATGTGACTAACCAGAGAG 2760  
QY 2760 ttccagtgatgaanaattgtgtccctcgtgtcagctgaagaacagctcgtgaacttcaaa 2819  
Db 2761 TTTCACTGATGCAAAATTTGTGTGCTCTGTGTTCACTGTAACAGTCTTGACTTTTAAA 2820  
QY 2820 accttgaataaagcttcccacagttgtataaatltgaaacttgaagaatttlaaactllag 2879  
Db 2821 AACTTGAATAAGCTCCACAGTGTATTAATTTAGCAATTTAGAAATTTAAACTTTAG 2880  
QY 2880 atgataatgtgttccatcttcaatcttcaatcttcaatcttcaatcttcaatcttcaat 2939  
Db 2881 ATGATCATTTTGTGCTCAATTTTATTTATTTTATTTTATTTTAAATGCAAAAGGACTTAA 2940  
QY 2940 atgaacttgaatctcgtttttaaagaattttaaagaacttgtatctatacatatgagc 2999  
Db 2941 ATGAACCTTGATCTCGTTTAAAGATTATTAATAAACATGTGTATCTATACATATGCG 3000  
QY 3000 tcttgaagactagccttcaactacatagagatatgatactcatalgtatgccatalaaac 3059  
Db 3001 TCTTGAAGACTTAGCTTTACACTACACTACAGATATGATCTCATATGATATCAATTAAC 3060  
QY 3060 ctgcagagtgatlttccagagtgctcgataactgttaataatcaatccatagagctgaaa 3119  
Db 3061 CTGACAGTGATTTTCCAGAGTGTGATGATGTTAATTAATTAATTAATTAATTAATTA 3120  
|||||







## AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeCarliano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Labrecque, K., Lamarez, R., Landers, T., Lechoczek, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Margulis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J., Menes, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

## TITLE

## JOURNAL

Submitted (18-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 22, 2000 this sequence version replaced g1:7582713.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

## Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)

## Project Information

Center project name: L9868

Center clone name: 746\_F\_15

## Summary Statistics

Sequencing vector: M13: M77815, 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 147035 bases at least Q40  
Consensus quality: 153931 bases at least Q30  
Consensus quality: 156692 bases at least Q20  
Insert size: 167000; agarose-fp  
Insert size: 158579; sum-of-ctrls  
Quality coverage: 4.0 in Q20 bases; agarose-fp  
Quality coverage: 4.2 in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently consists of 25 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

* 1 1728: contig of 1728 bp in length
* 1729 1828: gap of 100 bp
* 1829 3355: contig of 1527 bp in length
* 3356 3455: gap of 100 bp
* 3456 6090: contig of 2635 bp in length
* 6091 6190: gap of 100 bp
* 6191 9106: contig of 2916 bp in length
* 9107 9206: gap of 100 bp
* 9207 12073: contig of 2867 bp in length
* 12074 12173: gap of 100 bp
* 12174 16121: contig of 3948 bp in length
* 16122 16221: gap of 100 bp
* 16222 19552: contig of 3331 bp in length
* 19553 19652: gap of 100 bp
* 19653 23309: contig of 3657 bp in length
* 23310 23409: gap of 100 bp
* 23410 27884: contig of 4475 bp in length

```

## FEATURES

## source

```

* 27985 27984: gap of 100 bp
* 27985 32689: contig of 4705 bp in length
* 32690 32789: gap of 100 bp
* 32790 36488: contig of 3699 bp in length
* 36489 36588: gap of 100 bp
* 36589 43152: contig of 6564 bp in length
* 43153 43252: gap of 100 bp
* 43253 46195: contig of 2943 bp in length
* 46196 46295: gap of 100 bp
* 46296 51596: contig of 5301 bp in length
* 51597 51696: gap of 100 bp
* 51697 56969: contig of 5273 bp in length
* 56970 57069: gap of 100 bp
* 57070 67007: contig of 9938 bp in length
* 67008 67107: gap of 100 bp
* 67108 76018: contig of 8911 bp in length
* 76019 76118: gap of 100 bp
* 76119 85029: contig of 8911 bp in length
* 85030 85129: gap of 100 bp
* 85130 92117: contig of 6988 bp in length
* 92118 92217: gap of 100 bp
* 92218 103464: contig of 11247 bp in length
* 103465 103564: gap of 100 bp
* 103565 111248: contig of 7684 bp in length
* 111249 111348: gap of 100 bp
* 111349 122466: contig of 11118 bp in length
* 122467 122566: gap of 100 bp
* 122567 132062: contig of 9496 bp in length
* 132063 132162: gap of 100 bp
* 132163 148654: contig of 16492 bp in length
* 148655 148754: gap of 100 bp
* 148755 160979: contig of 12225 bp in length.

1. 160979
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-746F15"
/clone_lib="RPCT-11 Human Male BAC"
1. 1728
/note="assembly_fragment"
1829. 3355
/note="assembly_fragment"
3456. 6090
/note="assembly_fragment"
6191. 9106
/note="assembly_fragment"
9207. 12073
/note="assembly_fragment"
12174. 16121
/note="assembly_fragment"
16222. 19552
/note="assembly_fragment"
19653. 23309
/note="assembly_fragment"
23410. 27884
/note="assembly_fragment"
27985. 32689
/note="assembly_fragment"
32790. 36488
/note="assembly_fragment"
36589. 43152
/note="assembly_fragment"
43253. 46195
/note="assembly_fragment"
46296. 51596
vector_side:right"
clone_end:Sp6
46296. 51596
/note="assembly_fragment"
51697. 56969
/note="assembly_fragment"
57070. 67007

```

```

Db 60502 ACATCTTTTCTAGCGACCTAACGAAACCTTTTCAGTGATGCAATCATCTGACCTCT 60443
Oy 2946 ggctcagctgaacacagctccctggaact-ccaacacctggaatgaagtcaccagctgta 3004
Db 60442 GGTTCAGCTGAAAAAGCCCTGACCTTCACGAAACCTTGATCAAAATTTCCACTTGA 60383
Oy 3005 taattgacaactttaggaatttaactttagatgcatgcttgcatttatttc 3064
Db 60382 CAAATGGCCATTATGAGAAATTTCAACTCA---ATCATTGGCTCTCTTTATATTTT 60327
Oy 3065 atttatttctgttaagca--aacaggaacttaagaaacttctctgttcttaag 3122
Db 60326 ATTTTATTTTGGTGTGTGAATGGGACCTAAACAATAAATCTTGATCTTTTAAAG 60267
Oy 3123 attataaacaacatggtatctatacatatgctcttgaggactgcttcaactaca 3182
Db 60266 ATTATTAACAACTTACGTATGATACATATAGCTCTTCACAGACCACAGCTTCACTATG 60207
Oy 3183 ctacaggatgatctcgcatagtacgcataataacctgcaagtgatttc-agaagtc 3241
Db 60206 CTGTAGGATATGCTCTGTGATGACATATGACATTCGAGGTTAATTTCTCTGATGTC 60147
Oy 3242 tcgatactgtaatacatc 3261
Db 60146 TTTATACCTCTATTTTCTCTC 60127

```

```

RESULT 14
LOCUS G30939 385 bp DNA STS 29-SEP-1998
DEFINITION DRES50 Human DRES Homo sapiens STS genomic clone 22368 3', sequence
tagged site.
ACCESSION G30939
VERSION G30939.1 GI:1724032
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 385)
AUTHORS Banfi,S., Borsani,G., Rossi,E., Bernard,L., Guffanti,A.,
Rubboli,F., Marchitelli,A., Giglio,S., Coluccia,E., Zollo,M.,
Zuffardi,O. and Ballabio,A.
Identification and mapping human cDNAs homologous to Drosophila
mutant genes through EST database searching
Nat. Genet. 13 (2), 167-174 (1996)
96225443

```

Contact: Ballabio Andrea  
 Telethon Institute of Genetics and Medicine, TIGEM  
 Telethon Institute of Genetics and Medicine, TIGEM  
 via Olgettina 58, 20132 Milano, Italy  
 Tel: +39-2-215601  
 Fax: +39-2-21560220  
 Email: dres@tigem.it  
 Primer A: GTAAACATGATGAGACAG  
 Primer B: GGGCTGAAAAAGATGACCTAC  
 STS size: 157  
 PCR Profile:

Presoak: 94 degrees C for 2 minutes  
 Denaturation: 94 degrees C for 45 seconds  
 Annealing: 57 degrees C for 45 seconds  
 Polymerization: 72 degrees C for 45 seconds  
 PCR cycles: 35  
 Thermal cycler: MJR PTC-200

Protocol:  
 Template: 25 ng  
 Primer: each 1 uM  
 dNTPs: each 220 uM  
 Tag Polymerase: 0.025 units/uL  
 Total Vol: 15 uL

## Buffer:

MgCl2: 1.5 mM  
 KCl: 50 mM  
 Tris-HCl: 10 mM  
 pH: 8.3

The 5' end of the cDNA clone 22368 is homologous to the staufen  
 Drosophila gene product. EST from which STS is derived: T89105.  
 Forward primer is 70-1233; Reverse primer is 70-1234.

## FEATURES

## source

1..385  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="20, 12.10 cR from GC110F11 (10d >3.0)"  
 /clone="22368"  
 /clone\_11b="Human DRES"

## STS

primer\_bind  
 primer\_bind  
 complement(165..185)

BASE COUNT 135 a 79 c 71 g 100 t  
 ORIGIN

Query Match 10.8%; Score 380.2; DB 75; Length 385;  
 Best Local Similarity 99.2%; Pred. No. 2,1e-78;  
 Matches 382; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 3069 ttatttggtaagcaacaggaacttaatgaacttgcctgtttaaagattatt 3128  
 Db 385 TTTATTTTGTATGCAACAGGACTTAATGACCTTGATCTGTGTTAAAGATTATT 326

Oy 3129 aaaaacatggtatctatacatatgctcttgaggactgcttcaactactaacg 3188  
 Db 325 AAAAAACATTTGTATCTATATGCTCTTGAGGACTTGTCTCACTACATCAG 266

Oy 3189 gatagatcctcatgtagcctataaactgcaagtgatttcocagtgctcgatc 3248  
 Db 265 GATATGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 206

Oy 3249 tattaattcatcctccatgaagctggaagaatgaactgaagcttctgtatacagctgtg 3308  
 Db 205 TGTATATTCATCTCATCTCATCTCATCTCATCTCATCTCATCTCATCTCATCTG 146

Oy 3309 ttgctttgattgttgtaactgtaacaggaagtggtgacatgagggctctggtgtgt 3368  
 Db 145 TTGCTTTGATGT 86

Oy 3369 ccgatggaacacctgtagcctcgagtgtaagtaagtaagtaagtaagtaagtaag 3428  
 Db 85 CCGTATGGAACCTGT 26

Oy 3429 ggaatttctcccatggaatgta 3453  
 Db 25 GGAATTTTCTCCCATGGAATGTA 1

```

RESULT 15
LOCUS AC055879 160979 bp DNA HTG 22-MAY-2000
DEFINITION Homo sapiens chromosome 15 clone RP11-746F15 map 15, WORKING DRAFT
SEQUENCE, 25 unordered pieces.
ACCESSION AC055879
VERSION AC055879.2 GI:8017940
KEYWORDS HTG; HTGS; PHASE1; HTGS; DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 160979)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 15, clone RP11-746F15
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 160979)

```





```
* 53039 59129: contig of 6091 bp in length
* 59130 59229: gap of 100 bp
* 59230 65896: contig of 6667 bp in length
* 65897 65996: gap of 100 bp
* 65997 89427: contig of 23431 bp in length
* 89428 89527: gap of 100 bp
* 89528 95785: contig of 6258 bp in length
* 95786 95885: gap of 100 bp
* 95886 102762: contig of 6877 bp in length
* 102763 102862: gap of 100 bp
* 102863 110899: contig of 8037 bp in length
* 110900 110999: gap of 100 bp
* 121002: contig of 10003 bp in length
* 121003 121102: gap of 100 bp
* 121103 133048: contig of 11946 bp in length
* 133049 133148: gap of 100 bp
* 133149 145755: contig of 12607 bp in length
* 145756 145855: gap of 100 bp
* 145856 165006: contig of 19151 bp in length
* 165007 165106: gap of 100 bp
* 165107 188243: contig of 23137 bp in length
* 188244 188343: gap of 100 bp
* 188344 194407: contig of 6064 bp in length.
Location/Qualifiers
1..194407
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19"
/clone="RP11-73711"
/clone_lib="RPC1-11 Human Male BAC"
1..1755
/note="assembly_fragment"
1856..3429
/note="assembly_fragment"
3530..5346
/note="assembly_fragment"
5447..7338
/note="assembly_fragment"
7439..10297
/note="assembly_fragment"
10398..13590
/note="assembly_fragment"
13691..16293
/note="assembly_fragment"
16394..19065
/note="assembly_fragment"
19166..22713
/note="assembly_fragment"
22814..27165
/note="assembly_fragment"
27266..32377
/note="assembly_fragment"
32478..36572
/note="assembly_fragment"
36673..41657
/note="assembly_fragment"
41758..47024
/note="assembly_fragment"
47125..52938
/note="assembly_fragment"
53039..59129
/note="assembly_fragment"
59230..65896
/note="assembly_fragment"
65997..89427
/note="assembly_fragment"
89528..95785
/note="assembly_fragment"
95886..102762
/note="assembly_fragment"
102863..110899
/note="assembly_fragment"
```

```
misc_feature 111000..121002
/note="assembly_fragment"
misc_feature 121103..133048
/note="assembly_fragment"
misc_feature 133149..145755
/note="assembly_fragment"
misc_feature 145856..165006
/note="assembly_fragment"
misc_feature 165107..188243
/note="assembly_fragment"
misc_feature 188344..194407
/note="assembly_fragment"
misc_feature 194407
vector.slide:right"
BASE COUNT 50993 a 43965 c 46383 g 50458 t 2608 others
ORIGIN
Query Match 15.8%; Score 553.2; DB 54; Length 194407;
Best Local Similarity 90.4%; Pred. No. 1.2e-118;
Matches 665; Conservative 0; Mismatches 48; Indels 23; Gaps 6;
QY 2041 gctgcgtgaacatcttaagtgtgtgtgtgtgtgtgaaccacaagaatgacagatgcca 2100
|||||
DB 3297 GCTGCACATGAACACCTTAAGTCTGCTGCTGAGTGGACCAACAAGATACAGATGCGCA 3238
QY 2101 agaaagaagaagaccatgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2160
|||||
DB 3237 AGAACGAGAAATGATCAATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3178
QY 2161 ccattataaaatcccaacatatactgaataaactgaactgtcttgaaatttgaaat 2220
|||||
DB 3177 CCATTATAAAATCCACACATATACACTGAATATTAACCTGCTTGAATTTGGAT 3118
QY 2221 ttctgatactccagtggtgcgagagacacggtgtgtgtgtgtgtgtgtgtgtgtgtgt 2280
|||||
DB 3117 TTTCTATACCTTACTAGTGCGCCGAGAG--ACGCTGTGTAAAGATGTGGCACACACAGG 3060
QY 2281 aagacaacagaacacagaagagcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2339
|||||
DB 3059 AACACACACG-----GGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3007
QY 2340 tgaatgccactcgtgtgacctgtgcggtcccttaagcaatagcagctgtgtgtgtgtgtgtgt 2399
|||||
DB 3006 TCACGCCACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2947
QY 2400 ggtgtgccacgacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2459
|||||
DB 2946 GGTGTGCCACGACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2887
QY 2460 tgttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2517
|||||
DB 2886 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2827
QY 2518 tggctcaactcagacacatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2577
|||||
DB 2826 TGGCTTA-----GTTGGGACAAACCTGTGACAGCCACACAGAGAGAGCCCTTCA 2776
QY 2578 ccggtcccaagagtaaaacacacagagaataaactgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2637
|||||
DB 2775 CCAGCCCCAGACGTAAAGACACCAAGAAATCAATGCTTCTTTCAGCATGTGACATTA 2716
QY 2638 ctcttcctagtgtgcagagccacacacacacacacacacacacacacacacacacacacac 2697
|||||
DB 2715 C--TTCTAGTGTGCCAGGCCCCACACACTCTGTGAGTGTGCCACACACACACACTGCTT 2658
QY 2698 tctctccacagtgatctgtatcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2757
|||||
DB 2657 TGTCTTTCACACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2598
QY 2758 tatataaaatttcat 2773
|||||
DB 2597 TATATAAAATTTTCTT 2582
```

Db	149889	TCGTGGGATTCATGCGATGAGCCACCCTGGCCGACGACATATATATAATTTTCATTGGA	1498303
Oy	2778	gaatttcctcaattgtatattagtttaaatagcaagatttggaaactgtctcgaactgact	2837
Db	149829	GAGTTCTCAACGTGATCTAATTTATATCGCAACATTTGGAAACTTTTCTAGACGTGACTT	1497707
Oy	2838	tatcaataacttaaacccagaagaatcatatccatcgtctatgctgttgaacttttaatt	2897
Db	149769	TATCATATATCTAACCAGCAAGATTCATATCCAGTGTATGTAGTATGACATTTTATTTT	1497101
Oy	2898	cattgaactaacccagagacagtttcaagtatgacaattgtgtccctctgttcaagctgaa	2957
Db	149709	CATTGACTATACCCGACGACATTTTCAGTATGTCAAATTAACGCGCCATGTGCTCAGCTGAA	1496509
Oy	2958	acaagtcctggagacttcaaaaaaccttgtaataagttccccaagttgtataaatggacaat	3017
Db	149649	ACAGTCCTGGGACTTTCAAAACCTTGATATAAGTCTCCACAGTGTATTAATTTGGACAAAT	1495900
Oy	3018	ttagaataattttaaacttagatgatacttbgcttccatttcaattcaattcaattcttg	3077
Db	149589	TTAGGAATTTTAAACTTATAGATGATCATTTTCTCCATTTTATTTTATTTTATTTTGTG	1495300
Oy	3078	lttaabycaaacagagact --- aatgaaactltgatctcgtttttaagaattttaaaa	3133
Db	149529	TTAATGCMAACAGAGACTTAAACAATGACATTTGATCTCCATTTTAAATATTCCTTAAGA	1494707
Oy	3134	acattggtgatactataataatgactctcttgaggacttagcttccaactacaagaatat	3193
Db	149469	AAATCATATGTGCTATACCTTAACGAGGCTCTTGAGAGACTTGTAGCTTTCACCTACACAGGATAT	1494101
Oy	3194	gattcccatgtgactacatatataaacctgcagaagtatttccaagagtctcgataactgtta	3253
Db	149409	GATCTCCGACATAGTTTCACATATAAACTGCAGACGTGATTTTCCAGAGTGTCTGATATCTATTA	1493500
Oy	3254	attacaatccataatgagctgnaaagaatgacactagtttctcgtatacaagctgtgtgct	3313
Db	149349	ATTACATCTCATTAAGGCTGAANAAGATGACCATATCTTCTTATATACAGATGTGGTCT	1492900
Oy	3314	tttgatgtgtgtgtactcgtataacagaagtgtgtgcactgggc	3356
Db	149289	TTTGATGTTGTGTTACTATACACGAAGTGTGTGCACGTGGGCC	149247

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
12	AC068845/c	Homo sapiens chromosome 19 clone RP11-73711 map 19, WORKING DRAFT SEQUENCE, 27 unordered pieces.	AC068845	AC068845.3	gi:9966312	HTG: HTGS_PHASE1; HTGS_DRAFT.	human.
							Homo sapiens
							Euteleostomi; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
							1 (bases 1 to 194407)
							Birren,B., Linton,L., Nussbaum,C. and Lander,E. Homo sapiens chromosome 19, clone RP11-73711 Unpublished
							2 (bases 1 to 194407)
							Birren,B., Linton,L., Nussbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Campoliano,A., Castle,A., Chopelny,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S., Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D., Gadaen,J., Gardyna,S., Grinde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Labocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Marguis,N., McCarthy,T., McKean,P., McCurk,A., McKernan,K., McSheeters,R., Melidilim,J., Menius,L., Mihova,T., Miranda,C., Mienga,V., Morrow.J.

TITLE  
JOURNAL  
COMMENT

Murphy, T., Naylor, C.H., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Olliver, T.M., Oliver, J., Peterson, K., Pierre, N.,  
 Pisanelli, Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange, Thomas, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Testfay, S., Theodore, J., Tirelli, A., Travers, M., Trigilio, J.,  
 Vassiliou, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., J.  
 Young, G., Zahroun, J., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (10-MAY-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Sep 4, 2000 this sequence version replaced gi:8247880.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996:1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
 -----  
 Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 -----  
 Project Information  
 Center project name: 737\_11  
 Center clone name: L7549  
 -----  
 Summary Statistics  
 Sequencing vector: M13; M77815: 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 181875 bases at least Q40  
 Consensus quality: 187842 bases at least Q30  
 Consensus quality: 190220 bases at least Q20  
 Insert size: 190000; agarose-fp  
 Insert size: 191807; sum-of-contigs  
 Quality coverage: 4.7 in Q20 bases; agarose-fp  
 Quality coverage: 4.7 in Q20 bases; sum-of-contigs  
 -----  
 NOTE: This is a 'working draft' sequence. It currently  
 consists of 27 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.  
 -----  
 1 1755: contig of 1755 bp in length  
 1756 1855: gap of 100 bp  
 1856 3429: contig of 1574 bp in length  
 3430 3529: gap of 100 bp  
 3530 5346: contig of 1817 bp in length  
 5347 5446: gap of 100 bp  
 5447 7338: contig of 1892 bp in length  
 7339 7438: gap of 100 bp  
 7439 10297: contig of 2859 bp in length  
 10298 10397: gap of 100 bp  
 10398 13590: contig of 3193 bp in length  
 13591 13690: gap of 100 bp  
 13691 16293: contig of 2603 bp in length  
 16294 16393: gap of 100 bp  
 16394 19065: contig of 2672 bp in length  
 19066 19165: gap of 100 bp  
 19166 22713: contig of 3548 bp in length  
 22714 22813: gap of 100 bp  
 22814 27165: contig of 4352 bp in length  
 27166 27265: gap of 100 bp  
 27266 32377: contig of 5112 bp in length  
 32378 32477: gap of 100 bp  
 32478 36572: contig of 4095 bp in length  
 36573 36672: gap of 100 bp  
 36673 41657: contig of 4985 bp in length  
 41658 41757: gap of 100 bp  
 41758 47024: contig of 5267 bp in length  
 47025 47124: gap of 100 bp  
 47125 52938: contig of 5814 bp in length  
 52939 53038: gap of 100 bp

LOCUS AC011492/c 157803 bp DNA HTG 04-MAY-2000  
DEFINITION Homo sapiens chromosome 19 clone CTR-187L3, WORKING DRAFT SEQUENCE

ACCESSION AC011492  
VERSION AC011492.5 GI:7690131  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 157803)  
DOE Joint Genome Institute.  
Sequencing of Human Chromosome 19  
Unpublished  
2 (bases 1 to 157803)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On May 4, 2000 this sequence version replaced gi:7458736.  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 136279, BC101302  
Center clone name: CIT978SRB\_187L3  
-----  
Summary Statistics  
Consensus quality: 152548 bases at least Q40  
Consensus quality: 153824 bases at least Q30  
Consensus quality: 154424 bases at least Q20  
Estimated insert size: 156750; agarose-fp estimation  
Quality coverage: 9.52 In Q20 bases; sum-of-confids estimation  
NOTE: This is a 'working draft' sequence. It currently  
consists of 8 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

\* \* \* \* \*

1            2374: contig of 2374 bp in length  
\*         2375      2474: gap of unknown length  
\*         2475      4521: contig of 2047 bp in length  
\*         4521      4621: gap of unknown length  
\*         4621      7566: contig of 2945 bp in length  
\*         7566      7667: gap of unknown length  
\*         7667      12083: contig of 4417 bp in length  
\*         12084     12183: gap of unknown length  
\*         12184     18288: contig of 6105 bp in length  
\*         18289     18388: gap of unknown length  
\*         18389     63650: contig of 45262 bp in length  
\*         63651     63750: gap of unknown length  
\*         63751     111373: contig of 47623 bp in length  
\*         111374    111473: gap of unknown length  
\*         111474    157803: contig of 46330 bp in length.

Location/Qualifiers  
1. 157803  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="19"  
/clone="CTR-187L3"  
/clone\_lib="Caltech human BAC library B"  
BASE COUNT 40983 a 38390 c 39640 g 38079 t 711 others  
ORIGIN

Query Match          20.1%; Score 705; DB 35; Length 157803;

		Best Local Similarity 72.5%; Pred. No. 3.3e-154;	
		Matches 1206; Conservative 0; Mismatches 90; Indels 367; Gaps 8;	
QY	2041	gctgcgctgaacatcttaaaagtgcgtgcttgagttgagcaacaagtagagatagcca	2100
Db	150889	GCTGCACAGACACCTTAAAGTGTGCTGCTGATGTGGACCAACAAGACAGAGATGGCA	150830
QY	2101	agaacaggaagaacgagcaacatgtctgtgtgtgtgaggggtgagctgaaccttttcggcattgaa	2160
Db	150829	AGAACAGGAAGTGTATCAATTCGTGTGTGGAGGTCTTAACCTTTTGTGGCCATGAA	150770
QY	2161	ccattataaataccacaacataatactgaataactgaactgctttgaaatcttgagat	2220
Db	150769	CCATTATTAATAATCCCAACATATTACACTGAGATATTTAAACTGCTTTGAAATTTGGAT	150710
QY	2221	ttctgataactcctcagctg9gcccagagagacacgcttggtlaaagatgtg9gcaagcaagg	2280
Db	150709	TTTCGATACCTCTAGTGGGCCGAGAG--ACGGTGGGTAAAGATGTGGCCAGCAGAGG	150652
QY	2281	aagaacacagaacacacaaagagagcgcgctgtgtgcc-gagctgagacgtgtcgtgtgtgtg	2339
Db	150651	AACACAAACAG-----GGGAGGTGGGCTGTGGCCGGCTGGACATGTGTGGGGTTGTGG	150599
QY	2340	tgataggccacgcgcttgacacctg9gctccctcagcaatgacagagctgcgctgtg9ggaagaag	2399
Db	150598	TGACGGCCACCTGGGTGACCTGGCTGTCCCTACGCAATACCAAGCTGCTTTTGGGAGAGAG	150539
QY	2400	gagctgcccagcagcagctgtgtctccg9gagacacagcaatcacacaccccttg9gcaactc9	2459
Db	150538	GGCTGCCACGACAGCTGGCTCTCTCGGGGACACAGCAATATCCAGCCCTCGGGACACTTCC	150479
QY	2460	tgattggtctcttttttccctgtg-tgaagaagaacg9gacagccctctcaagc	2517
Db	150478	TGTTTGCTTTTTTTTCCTCCGTGGAAGAGAACTGAAGCATGAGACCTCTTCTCAAGC	150419
QY	2518	tggtctactcagacacatctgg9gaacaaacccctggacagcaatg9gag9ag9gctttga	2577
Db	150418	TGGCTTAG-----TTGGGACAAACCTGTGATGACGACACACGAGAGAGGCTTCAA	150368
QY	2578	cc9gcccagagctaaagacacc9aagaanaaatacaatgcttccctaagc9tgagccaa	2637
Db	150367	CCAGGCCCAAGACTAAAGCACACAGAGAAGAAAGTCAATGCTTCTTTTCAGCATGACCTAA	150308
QY	2638	ctttcttagtgtgcagcg9cccacacccctccctgcagtaglaacccaacaaatcaacactgct	2697
Db	150397	C--TTCTAGTGTGGCCACGGCCACACCACTTCTCGAGTGGCCACACCACTACCACTGCTT	150250
QY	2698	tctcttccaacagtgatctgtatattctgtattgaattatcttctttgttgatattctt	2752
Db	150249	TCTCTTCCAAACAGTATGTATTTCTTGTAGTTTCATTAATTTCTTTTATTTGATGACAC	150190
QY	2753	-----	2752
Db	150189	TATATAAATTTTCTCCTCTCCTGCTTCCTTCTTCTTTCTTTCTTTCTTTCTTTCTTCT	150130
QY	2753	-----	2752
Db	150129	TTCTTTCTTTCAGACAAGTCTACCTCTGTACCCAGCAAGCTGAGTGCACAGCGACAGTC	150070
QY	2753	-----	2752
Db	150069	TTTGGTCCCTGCACACTTGTGCTCTCCAGGTTTCAAGCAATTGTCTGCCTCAGCTTCCGAG	150010
QY	2753	-----	2752
Db	150009	TAGCTGGGCGACACACCAGCTAGTTTTTTTGTACTTTTAATGAAACGGGGTTTACCAT	149950
QY	2753	-----	2752
Db	149949	CTTGGCCAGAGCGTGTGTAAGTCTGTGACCTGTGATCCACCGCGCTCGGCTCCCAAG	149890
QY	2753	-----gaaactataaaatttcaattga	2777

```

/db_xref="taxon:10116"
/tissue_type="brain"
/dev_stage="adult"
1..3059
/gene="Stau"
227..1714
/note="RNA-binding protein; similar to Drosophila stau1;
contains four RNA-binding domains; rStau16"
/codon_start=1
/product="Staufen isoform Stau16"
/protein_id="AF98119.1"
/db_xref="GI:9754866"
/translation="MYKVPDPSRMOSTYSYGRGAGYPRYPFPVPLLYOVELS
VGGQFNGKGMPPVKHDPATARALRTLOSEPRLPVNGRESEENLSEISOVE
EIALKRLNLYNPFSPFLTOVARRESGPHKNEFTRXSVGEFEGSKSKIKKNA
RAYEOLRLPLPAYERVPRITKKSQPLCKIQTAADGQNPISRLAQIOQAKKE
KEPEYMLTERGLPRRREFVMQVGHHTAEAGSTKKVAKRNAANMLETIGFKVPQ
ADPAKPAKSEKTPVKKPDGRRKVFEPSPDEGNTSNKEDFPMYPLSHQPLAG
ILPMPEVAQAVSOGHHTKDEPARAAPNKAATVMAIARELILYGTSTABTILKS
MISSGVHPGPRTRPSEOLYLSRAQGFQVEYKDPKNNKNECVSLINCSOPLVLSH
GICKOVESCHDMAALNIKLISLIDQOSTEMPTGNGPSACGRC"
3043..3048
polya_signal
/gene="Stau"
BASE COUNT 817 a 778 c 749 g 713 t 2 others
ORIGIN

Query Match 35.88; Score 1254.8; DB 74; Length 3059;
Best Local Similarity 86.3%; Pred. No. 3,7e-283;
Matches 1429; Conservative 0; Mismatches 207; Indels 20; Gaps 3;

611 cagaagcataaccaccactgtagaactaaatgacatgctgctgacatgaaacac 670
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
165 CAGAAATATAGCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 224
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
671 caatgataaagcctgtgacccctactcgcagtcagtcacccataactacaacta 730
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
225 CAATGTATAGCCCGGTGAGCCCACTCGGATGCAGTCCACTACAGTATGCAATGC 284
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
731 gaggaagtgctatcccccagagtgactttacccaattccaagttcaacttact 790
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
285 GTGAGAGTGCCTATCCCAAGATCTTCACTTCACTTCACTTCACTTCACTTCACT 344
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
791 aagtgagaacttcgtgagagagagaatcttaaggcaagagaaacaagaagctg 850
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
345 AAGTTGAGCTCTCCGTGGAGAGACAGATTATGAGAGGAAATGAGACCACTTCG 404
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
851 cgaacaacaga tgcctgcgaagaagcgtctgagatccctgcagaatgagccctgcagaga 910
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
405 TGAACATGATGCCACATGCCGAGCCCTGAGACTCTGCAAGTGCAGCCCTTACCA 464
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
911 ggcctgagagtgaaatgagaagaatcgagaagaataatccaataaactgaaata 970
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
465 GGCATGAGCTAATGGAAGAAATCAAGAAAGAAACCTTAATAAATCAGAAATAAGCC 524
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
971 aagtgattgagatgacacttaaacgaactgctgtgaaatttcg----- 1015
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
525 AAGTGTTGAGATTGCCCTGACGCGAATTTGCTGTGAATTTTGAAGTCTTTTCCCTTGA 584
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1016 ---agtgagccggagagtgagccacccacatgaaagaactgtgacaaagtttcg 1072
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
585 CACAGTGCGCCCGGAGAGTGGCCCGCACACATGAGAACTTTGTACCAAGGTTTCAG 644
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1073 ttggagagtttggtgagagagtgaaagaaagcaagaatctcaagaataatgacg 1132
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
645 TTGGGAGTGTGTGAGCGAAGAGAAAGCAAGAAAGATCTCCAAAGAAAGATGCGG 704
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1133 ccataagctgtcttgagaagactgaagaagttaccgccccgcctgcagcttgaaagata 1192
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
705 CAAGGCGCGTGTGAGACCAAGAGGCTGCACCTTCACTGCTGTGAGAGCAATGA 764
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1193 agcctagatcaaaaaaagaacacccatagtcagagccagacagccagaatatg 1252

```

```

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 765 AGCCCAATATCAAGAGAAAAAGTCAAGCCCACTCAAGCTACAGACTGCCCCGACTATG 824
1253 gccagggatcaatccgattagccgactggtcccgatcccgagcgcaaaaaagagaag 1312
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 825 GCCAAGGATGAATCCATTATGACGACTGCGACGATTCAGACAGGCAAAAAAGAGAGG 884
1313 agccagagtaacgctctccacagagcagagccctcccgcccgagagagtttgatg 1372
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 885 AGCCGAGTACATCTCTCTTACAGAACGAGGTCTTCCACCTCGCAGGAGCTTTGTATGC 944
1373 aagtgaaagttggaacacacactgcaagaagaaacgagcgcccaacaagaagtgagcaag 1432
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 945 AGTGGAAGTTGGGACTATCACACCSCCAAGAGAGCAGGTACCAATAAAGATGCCAAGC 1004
1433 gcaatgacgcgagaaacatgctgtgagatccttggtttccaaagtcgcgacgagccca 1492
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1005 GCAATGCTGTGAGAACATCTCTGAGATCCTGGGGTTCAAAAGTCCCGAGCGCAGCCTG 1064
1493 ccaaacccgacatcaagtcagagagagagacacccataaagaacacagggagatgaga 1552
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1065 CCAAGCCAGCACTCAATTCAGAGAGAGACTCCAGTAAAGAAACAGAGATGGAAGCA 1124
1553 aagtaaccttttgaaacctggtctgaggatgaaatgggactagtaataaagaagatg 1612
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1125 AAGTAACATTTTGTGAACCTAGCCCTGGGATGAATAATGCAACCAATTAAGAGATG 1184
1613 agttcaagatgactatcatatgaatcaagcagctgctgctggaattctccatggtgc 1672
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1185 AGTTGAGATGCTCCCTATTAGCTATGACGATGCGCAGCGGAGTCTCCCTGATGTC 1244
1673 ccgagtgccgacgagcgtgtgagatgagcaagagacatcaccaagaatttcccgag 1732
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1245 CAGAGGTGCGCCAGCGTGTGGGTATGTCAGAGACACATACCAAGATTTGCCCAAGG 1304
1733 caagtcgaaatcctgtccaaagccacggttaactgcatgataagccgagatgtgtgacg 1792
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1305 CAGCTCCAAATCTGCGCAAGGCCACGAGTAACTGCAATGATAGCCGAGAAATTTGTGACG 1364
1793 ggggcaactgcgcaagccgagagccatlttaagaataacatctcttcagcgacgtac 1852
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1365 GGGGCACTCGCCCAAGCCGAGACATTTTAAAGATTAATCTCTTCAGGACATGATC 1424
1853 ccaatgacatcctcaagagacccctctgagcaactggaactctctccagagtcagagat 1912
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1425 CCAATGAGCTCGACAAAGACCTCTGACGACACTGTACTACTTCCAGAGCCAGGAT 1484
1913 tccagttgataaagaagacttcccaaaaaacaagaacgaattgtatctctatca 1972
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1485 TCCAGTTGATACAAAGACTCTCCCAAGAAACAACAAGAACGAGTGTATCTCTCATCA 1544
1973 atgtctccttcagcgaactctgacatgcatgcatgcatgcatgcatgcatgcatgcatg 2032
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1545 ACTGCTTTTACACACCTCTCTGACGCTGATGCGCAAGATGAGATGAGATGATGTC 1604
2033 atgatatgctgctgctgacatccttaagttgtgctgtgagttgagcaacaagaatgacg 2092
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1605 ATGATATGCTGCTGACATCAATCTTAAAGCTGCTGTCTGACTTGACCAACAGAGCAG 1664
2093 agatgcaagaacagaaagagacaaatgctgtgtgtgtgagagtgctgaaacttctcg 2152
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1665 AGATGCCAAGACAGAAAGAGACCAATTTACAGCTGCGGAGAGTGTGAACTATTTCTG 2124
2153 gcaatgaaacatataaatacccaacatataactgaaataactgaa-aactgtcttgaaa 2211
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1725 GCAATGAAACCATTTATAAAA-CCCAACATATATACCGAAATACTGAGAACTGCTTTGAAA 1783
2212 atttgaattctgatatccacagttgagcgagaga 2247
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1784 ATTTGAAATTTCTGATTAATCTCACTGAGGCGCAAGACCA 1819

```

RESULT 11

Query Match	35.8%;	Score 1256;	DB 74;	Length 2860;
Best Local Similarity	86.7%;	Pred. No. 1.9e-283;		
Matches 1420;	Conservative	0;	Mismatches 210;	Indels 8; Gaps 3;

Oy	611	cagaagacataaccctactcgttgaacctaataatgcactctgtcatgaactcttggaaaaaac	670
Db	264	CAGAAATATATAGCTTCTTACCATTTGACCTTAATGCATCTGTGTGAACTGGAAAAAGAAAAC	323
Oy	671	caatgtaaaagccgtgtgacccttactctcgaatgcagtcacaccataactacaacatga	730
Db	324	CAATGTTAAGCCCGTGGACCCCTCACTCTCGGATGTGATGCACCTACAGCTATGGCAATGC	383
Oy	731	gaggagtgcttatccccccgagttacttaccattccagttccagttccacttactatc	790
Db	384	GTGAGAGTGGCTTATCCCCCGACGATACTTTTACCATTTCATCTCCACCTTTACTCTAC	443
Oy	791	aagtgagacattctctgttggagaggaacaaatlaahtggsaaaggaagaagaagcagtg	850
Db	444	AAGTTGAGCTCTCCGTGGGGGACACAGCTTTAATGGGAAAGAAATGAGACCAACCCG	503
Oy	851	cgaacaacaga tgcctgtctcccaaaagcgtltgagatcctcgaatagtacgcccctgcgaaga	910
Db	504	TGAATACAGATGCCCCCTGGCCCGTGGCTGTGAGACTGTGAGATGAACCCCTGCCAGAAA	563
Oy	911	ggtctgaggtgtgaatlgagaagaaatccgaagaagaaatltcaataatctgaataagtc	970
Db	564	GGTGTGAGGTGAATGTGAAGAGCAGAGAGAAAGAAACCTCAATTAATCGGAAATTAACCC	623
Oy	971	aagttcttgatgttgcactttaaaccgaagacttgccttgatatttgaggttgcccgagaga	1030
Db	624	AAGCTTTGAAATTTGCCCTGACGGGAATTTGCTCTGTGAAATTTTGTAGTGCCCGCGGAGA	683
Oy	1031	gtggccaccaccacatgaagaactctgtgaccaagtttcglttgggtgagttgtggygg	1090
Db	684	GTGCCCCACACACATGAAMAATTTGTGACAGAGGTTTCACTTGGGAAATTTGAGGGG	743
Oy	1091	aagttgaagggaagaagcaagaatlttcaagaaaaaatgcgccatagctgttctltgag	1150
Db	744	AAGAGAGAGGGAAAAAGCMAAAGATCTCCACAGAGAAATCGGCCAGGCTGTCTGTGAGAGC	803
Oy	1151	agctggaagaagttacccgcccctgcctgcgaagtgtgaagatlaagactatgaatcaaaaag	1210
Db	804	AGCTTAGGAGAGCTGCCACCCCTCCTCTGTGTGAGAGGATGAAGCCCAATCAAGAAAGA	863
Oy	1211	aaacaaaaccatagttcaagccacagacaagaccccaatatatgcccaggygatacattccga	1270
Db	864	AAATCTACGCCACCTGCA-----AGACAGCCCGGATTTATGGCCAAAGGATGATCTTA	917
Oy	1271	tttagcgacttggcccagatattcaacagcagcaaaaaaaggaagaagaccagatataacgctcc	1330
Db	918	TTTATTAACCTTTCACACATATCCACAGGCAAAAAAAGAGAGGACAGATCAATGATCTCC	977
Oy	1331	tcacagaagcagagcctcccgccgcgcgaaggagtgtga tgcaggtgaaggttgaaac	1390
Db	978	TTTACAGAACGAGGCTTTCACAGTTCGACGAGGAGTTTGTATCGAGGTAAAGGTGGGCA	1037
Oy	1391	acactgcagaagaagcgggacccaacaagaagttgccaagcgcaatgcagccgagaaca	1450
Db	1038	ACACTGCAGAGAGGTGGTATCCCAATTAAGAGGTGGCCCAAGCGTAAATCGTGTGAGACA	1097
Oy	1451	tgcctggagatctcgtgtttaaagttcccgagcgagagcccaaaccccgacacaa	1510
Db	1098	TGCTCGAGATCTCTGGGCTTAAAGATTCCCCAGCGGAGGCTCTGCCAAGCCAGACACTCAAT	1157
Oy	1511	cagaggaagaagacccaataaagaacccaggygattgaaagaagaatlaaccttlttgac	1570
Db	1158	CAGAGAGAAAGACTCTCCGTTAAAGAAACAGAGACAGGAGAAAGTAAAGCTTTTGAAC	1217
Oy	1571	ctgagctctgggagatgaanaatvggagactagatataaagaatgaagttcagatagctctatc	1630
Db	1218	CTAAGCCCTGGGGATGAANAATGGAATCTATGTAACAGAGACAGAGATTCAGAGATGCTCTTTC	1277

[illegible]



```
|||||
Db 34725 CCTACGCAATAGACAGTCTGTGGGAAAGAGGCTGCCAGCAGCTGTTCTCCG 34666
Oy 2426 ggaacccagcagatccacaacctgggaacccctccggttcttcttccctcgt 2485
Db 34665 GGACACCCAGCAGATCCACACCTGGGACCTCCGTTTGGCTTTTTCCTCCCTGAT 34606
Oy 2486 gaaaagaacacagcagcagccctctcaagctggctcactcaagacacattgggaacac 2545
Db 34605 GAAAGAAAGAAAGGACACACCCCTTCTCAAGCTGCTCAGCAGACATTGGGACAAAC 34546
Oy 2546 cctggacagcagatggcagagagagcctttgacggcccccagagctaaagcacaagaga 2605
Db 34545 CCTGGACAGCAGCATGGCAGAGAGAGCCTTTTGACCGGCCCAAGAGCTAAAGACACAGAGA 34486
Oy 2606 aaatcaaatgtctcctactactcagcgtgacccaacttcttagtggccacggcccaaac 2665
Db 34485 AAATCAATGTCTCTCTACTCTGACCGGACCAACTTTCTAGTGTCCACGGGCCACACAC 34426
Oy 2666 ctccctgagtaaccacacacacacacacacacacacacacacacacacacacacacac 2725
Db 34425 CTCCGCGAGTACCCACACACACACACACACACACACACACACACACACACACACACAC 34366
Oy 2726 gttcaatatttcttcttggatgatalagacacataataaatttcaattggaaattct 2785
Db 34365 GTTTCATTATTTTCTTTGATTGATATGACACTATATAAATTTTCATTGGAATTTCT 34306
Oy 2786 caattgtatctagttaaatagacagcttggaaactgtctgagagctgaattacata 2845
Db 34305 CAATTGTATCTAGTAAATAGACAGTTGGAAACTTGTCTGAGACTGACCTTATCAATA 34246
Oy 2846 atctaaccgacaagaatataatccatgltglatggttagacatttatttcatctgact 2905
Db 34245 ATCTAACGACAAATATATATCCATGTGTATGTGTAGACATTTTATTTCATTGACT 34186
Oy 2906 aaccacagacaglttcagtgatgacaattgltgcccctcgttcaagtgaacacagtcct 2965
Db 34185 AACCCAGACAGTTCAGTGAATGCAAAATTTGTGCTGCTGTTGAGTGAACAGTCT 34126
Oy 2966 ggaacttcaaaaaacttgaataagctcccaagctgtatataaattggacaattggaaat 3025
Db 34125 GGACTTTCAAAAACTTGAATAGTCTCCACAGTTGTAATAAATTTGACAAATTTAGGAAT 34066
Oy 3026 ttaaaccttaagatgacattggttcacatttatttcaatttatttatttatttattt 3085
Db 34065 TTTAACTTTAGATATATCTTTGGTTCATTTTTATTTCATTTTTATTTTTATTATGCA 34006
Oy 3086 aacaggaacttaaatgaacttltgactcgttlttaaagalttaaanaaacattgtglatc 3145
Db 34005 AACAGAGACTTAAATGAACCTTGTGATCTGTTTTAAAGATTATTAANAACATTTGTATTC 33946
Oy 3146 tatacatatgctcttggagacttgaacttcaactacacagatagatccatgata 3205
Db 33945 TATATATATGTGCTTGTGAGGACTTGTGCTTCTACATACAGATATGATCTCATATGA 33886
Oy 3206 gtccatataaacctgcagatgaatttccagagtgctcgataccttataatcatctcca 3265
Db 33885 GTCCATATAAACCTGCAGAGTATTTTCCAGAGTCTCGATATCTTTAATTACATCTCCA 33826
Oy 3266 ttaaggctgaaagaatgaactacglttctgtatacagcgtgtgtccttltgatttgg 3325
Db 33825 TTAGGGCTGAAAGAAAGATGACCTGCTTCTGTATACAGCTGTGTGCTTTTGTATTTGG 33766
Oy 3326 ttaactgtacacagaagtgtgtgacactgagctcgtgtgtgtgtccgtatgaaacctgg 3385
Db 33765 TTACTGTACACAGAAAGTGTGTGACTGAGGCTCTCGTGTGCTGTGTGAAACCTGG 33706
Oy 3386 taagcctgcagattgaagctcctcactcaatttgaagcttgaaatttctcccat 3445
Db 33705 TAGCCCTGCGACTTAAGTACTGCTTCCATTTCTTTTAAAGCTGGAATTTTCTCCCAT 33646
Oy 3446 ggaatgtaagtaaaaacttaagtgttgcatacaataatgtaactaa 3496
|||||
```

```
Db 33645 GGAATGTAAGTAAACCTTAAGTGTGTCATCAATAAATGTAATCTAA 33595
RESULT 8
AF290989
LOCUS 3041 bp mRNA
DEFINITION Rattus norvegicus stauven isoform Stau-16 (Stau) mRNA, complete cds, alternatively spliced.
ACCESSION AF290989
VERSION AF290989.1 GI:9755162
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 3041)
AUTHORS Monshausen, M., Putz, U., Rehbein, M., Schweizer, M., Desgroseillers, L., Kuhl, D., Richter, D. and Kindler, S.
TITLE Two rat brain stauven isoforms differentially bind RNA
JOURNAL Unpublished
AUTHORS Monshausen, M. and Kindler, S.
TITLE Direct Submission
JOURNAL Submitted (28-JUL-2000) Institute for Cell Biochemistry and Clinical Neurobiology, University of Hamburg, University Hospital Eppendorf, Sudefeldstr. 24, Hamburg D-22529, Germany
FEATURES
Source
Location/Qualifiers
1..3041
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue-type="brain"
/dev_stage="adult"
1..3041
/gene="Stau"
227..1696
/gene="Stau"
/note="RNA-binding protein; similar to Drosophila stauven; contains four RNA-binding domains; rStau-16"
/codon_start=1
/product="stauven isoform Stau-16"
/protein_id="AAF98177.1"
/db_xref="GI:9755163"
/translacion="MYKVPDHSRMOSTYSYGMRGAYPPRYFPVPPVLLYQVELS VGGQFNKGKMRPPVKKHDTARARLRTLOSPPLEPERLEVNGRESEENLNKSLISOYF EIALKRLNLPVNEVARESGPEHMKNFVTRVSVGEFVGBGSKSKISKNAARAVLSD LRLRPLPVADEVPRIRIKKSQPTCKLOTAPDYQGNPNISRLQIOAKKEKPEEM LTERGLPRRREPVMOYVGHHTAEGAGTNKKVAKRNAEEMLEILGFKVQOAPAPK ALKSEKTPYVKPDGKRKVPFEESPDENGTSKDEDFRMPYLSHOOLPGILPVPV EVAQVGVSGHHKDKFRAAPNPAKATVAMLRLEILLGGTSPTAETILKNSISGH VPHGPRTRPSPOLYLLSRAQGFQVEYEDFRFNKNKNECVSLNCSOPPLVSHGKIDV ESCHDMAALNLTKLLELSDOSTMPRTGNGVPSACGRC"
3025..3030
/gene="Stau"
BASE COUNT 814 a 771 c 747 g 707 t 2 others
ORIGIN
Query Match 36.6%; Score 1282.8; DB 74; Length 3041;
Best Local Similarity 87.2%; Pred. No. 1e-289;
Matches 1429; Conservative 0; Mismatches 207; Indels 2; Gaps 2;
```

```

Db 2037 CGGCCCCACCACTCCTGCAGTACCCACACATCCACTGCTTTCTCTTCCACAGCTGA 2096
Qy 2714 tctgtatctctagtttcattatcttcttgatgta latagacataataaatttcac 2773
Db 2097 TCGTATTTCTTACTTTCATATTTCTTTTGATTGATGATGACACTATATATAATTTTTCAT 2156
Qy 2774 ttgagaattctcaattgtatctagtttaaatagacagtttggaacctgtctgagctg 2833
Db 2157 TTGAGAAATTTCTCAATTTGATCTAGTTAAATACACAGTTTGGAACCTGTGTGACACTG 2216
Qy 2834 acttctacataatctcaacgaagaagaatcatcatctgtatctgtgttagacatttt 2893
Db 2217 ACTTTTCAATATATCTAACCGACAAAGATCATATCCATGTATGTGTAGACATTTT 2276
Qy 2894 attctatctactaaccaagagacagttctcagtgatgcaaatgtgtccctctgtctcagc 2953
Db 2277 ATTTCATTTGACTTACCCAGACAGTTTCAAGTATGCAAAATGTGTGCCCTCTGTGTTACG 2336
Qy 2954 tgaacagctcctgagcttccaaaacctggaataagctctccacagttglatataactgga 3013
Db 2337 TGAACAGCTCCTGAGCTTCAAAAACCTGGAATAGCTTCCACAGTGTATTAATTTGGA 2396
Qy 3014 caattaggaatttaacttagatgacatcttggttcatttcttattctattt 3073
Db 2397 CAATTTAGGAATTTTAACTTTAGATGATCATTTGTGCTCCATTTTATTTTCAATTTTAT 2456
Qy 3074 ttgttaataagcaacagagacttaaatgaaacttgatctcgtttttaaagaatttaaaa 3133
Db 2457 TTTGTTATGCAACAGAGACTTAAATGAACTTGTGATCTCTGTTTAAAGATTTATTAATA 2516
Qy 3134 acattgttatctatatacatatgagctcttgagagacttgcttcaactacatacagatat 3193
Db 2517 ACATGTGTATCTATACATATGAGCTCTTGAGACTTGTGTTTCACTACACTACAGATAT 2576
Qy 3194 gactctcagtgatgctcatataaaccctcgagagtgatcttccagagtgctcagatcgtta 3253
Db 2577 GACTCTCAGTGTACTCATATATAAAGCTGACAGATGATTTTCCAAAGTCTTCGATACCTGTTA 2636
Qy 3254 attacatctccatctagagctgaaagaagaalpacctacgtttctgtatatacagctgttgc 3313
Db 2637 ATTACATCTCCATTTAGGCGCTGAAAGAAATGACCTAGCTTTCTGTATACAGCTGTGTTGCT 2696
Qy 3314 ttgtatgtgtgttactgtatctgacacagaagtgctgacacagagctctgcgtgtgttcgta 3373
Db 2697 TTGATGTGTGTGTACTGTACACAGAAAGTGTGTGACTGAGGCTGTGCGGTGTGCTCCTGA 2756
Qy 3374 tgaagaacctgtgagccctcgagagtgtaagtactgcttccatctatgtttagcgtggaat 3433
Db 2757 TGAAGAAGCTGTGAGCCCTGCGAGTTAAGTACGTTCATTTGTTTACCTGTGAAT 2816
Qy 3434 ttctctcccatggaatgttaagtaaaacttaagtgttctgcatcaataaagtgtaact 3493
Db 2817 TTTTCTCCCATGGAATGTAGTAATAACTTAAGTGTGTGTGATCATCAATAATGTGTAATACT 2876
Qy 3494 aaa 3496
Db 2877 AAA 2879

```

```

RESULT 7
AL133174/c AL133174 113836 bp DNA PRI 14-JUN-2000
LOCUS Human DNA sequence from clone RP3-470L14 on chromosome 20, complete
DEFINITION sequence.
ACCESSION AL133174
VERSION AL133174.15 GI:8573761
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 113836)
AUTHORS Ramsay, H.

```

TITLE Direct Submission  
JOURNAL Submitted (14-JUN-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
COMMENT On Jun 20, 2000 this sequence version replaced gi:8388434.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; Sw.; SWISSPROT; Tr.; TrEMBL; Wp.; WormPep; Information on the WormPep database can be found at <http://www.sanger.ac.uk/Projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone configs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>

RP3-470L14 is from the library RPl-3 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>

VECTOR: pCIPAC2

IMPORTANT: This sequence is not the entire insert of clone RP3-470L14. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true right end of clone RP3-470L14 is at 113836 in this sequence. The true right end of clone RPl-15566 is at 100 in this sequence.

## FEATURES

Location/Qualifiers

1..113836

source

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="20"

/clone="RP3-470L14"

/clone\_11b="RPl-3"

/clone\_11b="RPl-3"

3061..3077

misc\_feature

/note="Weak data"

95583..95627

misc\_feature

/note="Single clone region. pUC subcloned to give 12x coverage. Assembly confirmed by restriction digest data."

BASE COUNT 33974 a 25115 c 24233 g 30514 t

ORIGIN

Query Match 38.7%; Score 1357.4; DB 34; Length 113836;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1369; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```

Qy 2127 gttgtggaagtggtgaacctttcttgccatgaacattataaattccaacataatac 2186
Db 34965 GTGTGGAGGTGCTGAACCTTTCTGCGCATGAACCAATTAATAATCCAACTATATATAC 34906
Qy 2187 tgaataactgaacacgcttgaataatttgaaattctgataacctcagtgagccgagag 2246
Db 34905 TGAATAACTGAACACTGCTTTGAAATTTTGAGATTTCTGATACCTCAGTGGCGCAGAG 34846
Qy 2247 acacggttggttaaggaatgtggcagcagcaggaagaacaagaagaagagcg 2306
Db 34845 ACACGCTGGGTAAAGCATGTGGCAGCAGCAGGAGACACAGAAACACAAAGAGCGCG 34786
Qy 2307 ctgtgccc-gcctgagactgtgctggggtttgtgtatgagccactgtgagctgcggt 2265
Db 34785 CTTGTGCGCGGCTGGACCTGTGCTGGGTTGTGTATGTGCTGACCTGTGCGGG 34726
Qy 2366 ccctacgcaatagcagctgctgtgtgggaagagcgctgccagcagctgtctcccg 2425

```



BASE COUNT	891 a	693 c	699 g	783 t
ORIGIN				

Query Match	79.8%;	Score 2799.4;	DB 68;	Length 3066;
Best Local Similarity	99.5%;	Pred. No. 0;		
Matches 2870;	Conservative 0;	Mismatches 6;	Indels 7;	Gaps 6;

OY	617	gcaataacccctactytagaactaaatgcaciytgcatagaacctvgnaaaaaaaccaalgt	676
Db	1	GCATTAACCCCTACTGTAAGAACTTAAATGGCACTGTGCATGAATACTTGAAAAAACCAATGT	60
OY	677	ataagcctgtctggacccttaactctcgatbcagtlccaacctatracacaacatlgagagag	736
Db	61	ATAAGCCTGTTGACCCTTACTCTCGATGCAGTCCACCTATTACTACAACATGAGAAGAG	120
OY	737	gtgccttatcccccgaggtaactltiltacccaatlccagatlccoaacttiactlilatcaatg	796
Db	121	GTCGTTATFCCCCCGAGGTACTTTTACCATTTCAGTTCCACCTTACTTTATCAAGTGG	180
OY	797	aacttcitfvggagaagacgaactlaattatgycaaaagaagaacaagacagctgcgaac	856
Db	181	AACTTTCTGTGGAGGAGCACCAATTTAATGGCAAAAGAAAGCAAGACAGCTGTGGAAAC	240
OY	857	acgaatgtctgcgccaaagagtltgagatcccttcagaaatbagcccctgcagaagagctcg	916
Db	241	ACGATGCTGCTGCCMAAGCTTGAAGATCTCTGACAGTATGAGCCCTTGCAGAGAGCTGG	300
OY	917	aggttgaaatggaaagaaatccgaaagaagaatatcatlaaatcltaaaataagtaacgt	976
Db	301	AGGTGAATGGAAAGAAATCCGAAGAAAGAAATCTCAATTAATCTGAATTAAGTCAAGT	360
OY	977	tltgagatltgcaacttaaacggaacttgcctgtgaattlttgaggttgcgggagaagttg	1036
Db	361	TTGAGATTGACATTAACCGSAACTTCTCTGTGAATTTCAAGGTGGCCGCGGAGAGATGGCC	420
OY	1037	cacccacatagaagaacttltgtaccaaggtlttcgvtbgggagcttgttggggagaaggt	1096
Db	421	CACCCACATGASAGAACTTGTGTACCAAGSTTTCGTTGGAGTGTGTGGGGCAAGGTG	480
OY	1097	aaggaaagaacaagaagatltaaagaanaatgpcgcatalgctyltcttgaagagctga	1156
Db	481	AAGGAAAAGCAAGAAATTTCAAAGAAAAATGCCCATACGTTCTTGAAGAGCTGA	540
OY	1157	agaagttacccggcccttgcgcgcagttgaaocggtlaaogcctaagatacaaaaagaanaaa	1216
Db	541	AGAAATTACCGCCCTCTGCCTGCAATTGAACGATGAAGCTTAGAATCAAAAAAGAAACAA	600
OY	1217	aaccatagtatcaagcccaacaagaagcccaagaataltgccaagggatlcacatccgatlla	1276
Db	601	AACCATATGTCAAGCCACACACAAAGCCCAAGATATGTGCCAGSGGATCAATCCGATTAGCC	660
OY	1277	gaetfgcccaagatlcagcaagcaaaaaagagaagagcacagatatacgcctcctcacag	1336
Db	661	GACTFGCCCAAGATCCACACAGCAAAAAAGAGAGAGCACAGATACACGCTCTCACAG	720
OY	1337	agcgagagcctcccgccgcgcgaaggaattgtgtatcgagtgaaggttbgaaaaaceactg	1396
Db	721	AGCGAGGCTCTCCGCGCCGACGGGAATTTGTATGTACGATGAAGGTTTGAAAAACCACTG	780
OY	1397	cagaagaagacgggacccaacaagaaggttggccaagcgcaatbcagccgagaaacatgctg	1456
Db	781	CAGAAGGACGGGCGACCAACAAGAAAGTGTGCCAAGCGCAATCAGCCGAGAAACATGCTGG	840
OY	1457	agatccttlygtlttcaaaagttcccgacgpgcagcccaacaacccgacatcaagtcagag	1516
Db	841	AGATCTCTTGTTTTCAAAAGTCCCACACG--CAGCCACCAAAACCCCACTCAAGTACAGG	898
OY	1517	agaagaacccaataaanaaaccaaggaatlggaagaaatlaac--tlttttgaacctg	1575
Db	899	AGAAAGACCCCATTAAGAAACACAGGGGATGGAAGAAATAACCTTTTTTTTGAACCTGGC	958
OY	1576	tcttggagat-gaaatbtggactagtaataagaagatgtgatttccagatgtccttactaa	1634

Db	959	TCGTGGGATGCAAAATGAGGACATAGTATAAAGAGATGATTCAGATGCCCTATCTAAG	1018
QY	1635	tcatacgaagctgcctcctcgtgaattcttcccatagtgcccgaggtgcgcgaagctgttag	1694
Db	1019	TCATCAGCAGCTGCCTCTCGGAATCTCTCCCATAGTGCCCGAGGTGCCACGAGCTGTAAG	1078
QY	1695	agttagtaagaagacatacaaccacaagaatttaccagggcagctccgaatccctgcgaagc	1754
Db	1079	AGTTAGTCAAGACATCATCACCAAAAGTTTACCAAGGAGCTCCGAATCTCGCCAAAGC	1138
QY	1755	caagytacctgcctatgaagcccgagagttgttgtatbgyygccacctcgccacagccga	1814
Db	1139	CACGGTACTCCCATGTATACCCGAGAGTGTGTGTGGGGGACCTCGCCACAGCCGA	1198
QY	1815	gaccattttaagaataaacatctcttcagggccgcgtaaccccatlgaacctctacgaagc	1874
Db	1199	GACCAATTTTAAAGATACATCTCTTCAGGCCACGTACCCCATGAGCCTCTACCAAGAC	1258
QY	1875	ctctbagaacctlgaacctcttccagagtcagaggtaccaggtctgaatacaagaactt	1934
Db	1259	CTCTGAGCACTGGACATATCTTTCACAGTCCAGGGATTCAGGTTGAATACAAAGACTT	1318
QY	1935	ccccaaaacaacaagaacgaatttgtatctctatcaaltgtcctcctcagccactct	1994
Db	1319	CCCCAAAACAACAAGAACGATTTGTATCTTTATCAATTCCTCCTCACCCACTCT	1378
QY	1995	gatacgcaatggtatcggcaagatgtggagctccgcatagtatagctgcgtctgaacat	2054
Db	1379	GATCAGCCATGTTGTCGGCAAGATGTGGAGTCTCTCCATGATGTGGCTCCCTAACAT	1438
QY	2055	cttaagctgtctcctgaatttgagccacaagaatacagaatgcacaagaacgagaacg	2114
Db	1439	CTTAAAGTTCCTGTCTGATGTTGAGCACCACAAACTACAGAGATGTCCAAAGACGAAAGG	1498
QY	2115	accaatgtctgtgtgtggaggtgtcctgaacctttctcggcaltgaacaltataaattcc	2174
Db	1499	ACCAATGCTGTGTGGTGGAGGTCTGAACCTTTCTGGCCATGAAACATTTATAAATCC	1558
QY	2175	caacataatacttgaataatactcgaacctgtcttgaataatttgaattctcgaacctca	2234
Db	1559	CACACTATATCTGAAAAATCTGAAATGCTTTTGAATAATTTGGAATTTCTGTACTCTCA	1618
QY	2235	gtgggcccagagacaggtgtggttaaagatgtgtggcagcagcaggaagaacaacgaac	2294
Db	1619	GTGGGCCAGAGAACGCTGGTGAAGATGTGGGCACGACGAGGAGAACACACGAAC	1678
QY	2295	acaaagagcgagctgtggcc-ggcctgagctgtcctggggttttgttgtgaatggccactcg	2353
Db	1679	ACAAGAGAGCGGCTGTGGCCGGCGGTGACATGTCGGGGGTTGTGTGATGGCACCTCGG	1738
QY	2354	tgacctgtgvggttccctcagcaatagcagctgtcgtgtgggaagaaggtctccacgacag	2413
Db	1739	TGACCTGGCGGTCTCCCTACGCATATAGCAGCTGCTGTGGGAAGAGGGGTGCCACGACAG	1798
QY	2414	ctgagcttcccgaggacaacccagatcccaaacccctbggagacctccggtgttgttcctttt	2473
Db	1799	CTGGTTCTCCGGGACACACAGAGATCCACCCTTGGCACTCCGTGTGTTTGCTTTT	1858
QY	2474	tttccctgtgtgaagaagaagaagcagaccctctcgaagctgtgctcaactcagacac	2533
Db	1859	TTTTCCCTGTGTGAAGAAGAAAGGACAGACACCCCTTTCAGCTGTGCTCACTCAGACAC	1918
QY	2534	attgggacaacaccttgacagccatctccagagagagagctcttgaacggtcccaagctcaa	2593
Db	1919	ATTGGACAATA-CCTTGACAGCCATGTCCACAGAGAGAG-CTTTTACCGGGCCCAAGAGCTAA	1976
QY	2594	aagccacagagaaaaataaatgtcttccactaagcgttgagcccaactttctaggtgtgca	2653
Db	1977	AAGCACGAGAAAAATATAATGCTTCTTACTCAGCGTGAACCAACTTTTCTAGTGTGCCA	2036
QY	2654	cggcccccacccctctgcagtaaccacacatacaacaaatgtctctctcacaacgtga	2713

[illegible]

Qy	3131	aaaacatggtgatactataaiaitggtctctttagagacttagcttccactaacacaaaga	3150
Db	2810	AAAACATTTGTGTCTATCTATACATATGGCTCTTGAGCACTTAGCTTCTACTACATACAGA	2859
Qy	3191	tatgatactccatgtaagtcacataaaacctgcagagtgatlltccagaagtgcgatactg	3250
Db	2870	TATGATCTCCATGTAGTACATATAAACCTGCAGAGTGATTTCCAGAGTGCATCTAGT	2929
Qy	3251	ttaattacatcccatlaagagctgaaagaatgacctagcttctcgtatacaagctggtc	3310
Db	2930	TTAATTACATCTCCATTTAGGGGCTGAAGAAATAAGTACCTAGCTTTCTGTATACAGCTGTGTT	2989
Qy	3311	gcttltgatitgtgttactgatacaagaagtgtgtgcactgaagctctgcgtgtgccc	3370
Db	2990	GCTTTTATGTGTGTGTACTGTACGTACAGAGAAGTGTGTGACTAGGCTCTGGCTGTGGTCC	3049
Qy	3371	gtatgaaaaacctgtagacctgcgagttaagttaactgcttccatcattgttlaacgtcg	3430
Db	3050	GTAATGGAAGCGTGTAGCCCTCGAGTTAAGTACTGCTTCATTCATTGTTTACGCGCG	3109
Qy	3431	aatttctccccaatgaaatgtaagtaaaactaaagtgttgttgcataaataatgtgtaat	3490
Db	3110	AATTTTCTCCCCATGGAATGTAAAGTAAACACTTAAGTGTTCATCAATAAATGTAAAT	3169
Qy	3491	actataaaaaaaaaa 3506	
Db	3170	ACTATAAAAAAAAAA 3195	

FEATURES	source
LOCUS	HSAL132258
DEFINITION	HSAL132258 3066 bp mRNA
ACCESSION	AJ132258
VERSION	AJ132258.1 GI:4572587
KEYWORDS	staufen gene; staufen protein.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 3066)
TITLE	Falcon,A.M., Fortes,P., Marion,R.M., Beloso,A. and Ortin,J.
JOURNAL	Interaction of influenza virus NS1 protein and the human homologue of Staufen in vivo and in vitro
MEDLINE	Nucleic Acids Res. 27 (11), 2241-2247 (1999)
REFERENCE	99263025
AUTHORS	2 (bases 1 to 3066)
TITLE	Ortin,J.
JOURNAL	Direct Submission
DEFINITION	Submitted (25-JUN-1999) Ortin J., Molecular and Cellular Biology, Centro Nacional de Biocetnologia, Cantoblanco, 28049 Madrid, SPAIN
LOCATION	Location/Qualifiers
1..3066	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
36..1526	
/gene="staufen"	
36..1526	
/gene="staufen"	
/function="dsRNA binding protein"	
/codon_start=1	
/product="staufen protein"	
/protein_id="CAB40082.1"	
/db_xref="GI:4572588"	
/translation="MKLGKPMYKPPVDYSRMSTYNNMRGAYPPRYYPVPPPL LYQVELSVGGOQFNGKGTROAAKHDAALRIQNEPLPERLEVNGRSEENLNK SEISQVEIALKRNLPVNEFAKSGPHHKNVYTKSVGEFVGEKGSKATISKKA AIALVEELKRLPLPAVERVKPRIKKTKPIVRQPSPEVGGINPLISRLAQIQAAK EKBEYETLLTERGGLPRRREVMGVNHTAEGGTNKVKAKNAAMNLEIGFVQ ORSNQTTOVGREDTHKETRGKMSNLEFLNADGEMGTSNKDEPFEMPLISHOOLPA GLPMEVEVAAQVYSGCHHTKQETRAAPKATYVAMTARELITGSCSPVETLIK NNISGDSHPGPIPTRESDLDYLSRVOGEFVEERKDPKNNKNEFVSLINCSQPLIS HGIGKDVDSOHMAALNLIKLLSLDDOSTEMPTGNGPMSVGCRC"	

ACCESSION AK025519  
 VERSION AK025519.1 GI:10438059  
 KEYWORDS Oligo capping, fis (full insert sequence).  
 SOURCE Homo sapiens hepatoma cell\_line:hepg2 CDNA to mRNA, clone\_lib:hep  
 clone:HEP02379.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (sites)  
 AUTHORS Kawabata,A., Hiki,J., Kobatake,N., Inagaki,H., Ikema,Y.,  
 Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,  
 Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.  
 NEBO human cDNA sequencing project  
 JOURNAL Unpublished (2000)  
 REFERENCE 2 (bases 1 to 3194)  
 AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,  
 Shibahara,T., Tanaka,T. and Nakamura,Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio  
 Sugano, Institute of Medical Science, University of Tokyo,  
 Laboratory of Genome Structure Analysis, Human Genome Center,  
 Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan  
 (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,  
 Fax:81-3-5449-5416)  
 COMMENT NEBO human cDNA sequencing project supported by Ministry of  
 International Trade and Industry of Japan; cDNA full insert  
 sequencing: Research Association for Biotechnology; cDNA library  
 construction, 5'- & 3'-end one pass sequencing: Department of  
 Virology and Human Genome Center, Institute of Medical Science,  
 University of Tokyo (partly supported by Science and Technology  
 Agency).  
 FEATURES  
 source Location/Qualifiers  
 1..3194  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_line="hepg2"  
 /cell\_type="hepatoma"  
 /clone="HEP02379"  
 /clone\_lib="HEP"  
 /note="cloning vector pME18SF13"  
 misc\_feature 1..3194  
 /note="highly similar to AF061938 Homo sapiens staufen  
 protein (STAU) mRNA"  
 BASE COUNT 894 a 787 c 732 g 781 t  
 ORIGIN  
 Query Match 81.3%; Score 2848.8; DB 34; Length 3194;  
 Best Local Similarity, 99.7%; Pred. No. 0;  
 Matches 2886; Conservative 0; Mismatches 7; Indels 3; Gaps 3;

Db 592 GCTGGAGTGAATGGAAGAAATCCGAAGAAATCTCATATATCTGAATAAGTCA 651  
 QY 972 agtgtttgagattgcacttaaaagaaacttcctgtgaattcagagtgccggagag 1031  
 Db 652 AGTGTGATGATGACCTTAACGGAACCTTGCCTGTGAATTTGAGGTGCGCCGGAGAG 711  
 QY 1032 tggccacccccacgaagaatttgtaaccaaggttcggttggggagttgttgggga 1091  
 Db 712 TGGCCACCCCACTATGAACACTTTGTGACCAAGTTTGGTTGGGGAAGTTGTGGGGA 771  
 QY 1092 aggtgaaggaaagcaagaagatttcaaaagaaatccgcacatagctgttcttgaga 1151  
 Db 772 AGCTGAAGGAAACAGAAATTTCAAGAAATATGGCGCATAGCTGTTCTTGAGGA 831  
 QY 1152 gctgaagaagtacccgcccctgcctgcagtgtaacgagtaaaagctagaatcaaaaga 1211  
 Db 832 GCTGAAGAAGTTACCGCCCTCTCGAGTTGAAGAGTAAGCTAGAACAAAGAA 891  
 QY 1212 aacaaaccatagtcacagccacagacaaagcccaatatatgcccaggagatccgat 1271  
 Db 892 AACAAAA-CCATATGTCAGCCACAGACAAAGCCCAATATGGCCAGGGATCAATCCGAT 950  
 QY 1272 tagccgactggccagatccagcagcagcaaaaggagaagagcagagtaacgctcct 1331  
 Db 951 TAGCCGACTGGCCAGATCCAGCAGGCAAAAAGAGAAAGAGCCAGACTACACGCTCCT 1010  
 QY 1332 cacagaagcgaagcctccgcgcgcgagggagttgtgatcagtgtaagtgtaaaaca 1391  
 Db 1011 CACAGAGCAGGCTCCCGCCGCGCAGGAGATTGTGATGACGAGTGAAGTTGAAACCA 1070  
 QY 1392 cactgcagaagaagcggcaccacaaagaagtggtgcacagcgaatgcagccgaagaat 1451  
 Db 1071 CACTGCAGAAAGAGGAGGACCAACAAAGAGGTGGCCAAAGTGCATATGCGCGAAGACT 1130  
 QY 1452 gctgagagatccttggtttcaaaagtcocgcagcagcagcccaaacccgcaactaagtc 1511  
 Db 1131 GCTGGAGATCCTTGGTTTAAAGTCCCGAGGCGCAGCCACCAAAACCGCAGCTCAAGTC 1190  
 QY 1512 agagagaagacacccataaagaacccagggagtggaagaaagtaaccttttgaacc 1571  
 Db 1191 AGAGGAGAAAGACCCCATTAAGAAACCA-GCGATGGAAGAAAGTAACCTTTTGAACCC 1249  
 QY 1572 tggcctggggatggaataatggagctagttaataagaagatggttcagatgcttaact 1631  
 Db 1250 TGGCTGTGGGATGAAAGAGGACTAGTAATTAAGAAGATGAGTTGAGGATGCTTATCT 1309  
 QY 1632 aagtcacagcagcctgcctgcgaattctcccatggtgcccgaagtcgcgcaagctgt 1691  
 Db 1310 AAGTCATCAGCAGCTGCTGCTGGAATTTCTCCATGATGCGCCGAGGTGCGCCAGGCTGT 1369  
 QY 1692 aggagttagtcaggacatcacacaaagaatttaccagggcagctcogaatccttgccaa 1751  
 Db 1370 AGGAGTTAOTCAAGACATCACCAAAAGATTTTACAGAGGCGAGCTCCGAATCTGCCAA 1429  
 QY 1752 ggcacagtgtaactgcagatagacccggagagttgttgtaaggggcagcttcgccaacgc 1811  
 Db 1430 GGCACGGTAAGTGCATGATATGCCCCGAGAGTTGTGATGGGGCACCTCCGCCACAC 1489  
 QY 1812 cgagaccatttaagaataacatccttcacagcagctaaccccatggaacctcgaag 1871  
 Db 1490 CGAGACATTTTAAGCAATTAACACTCTTCAGGCCAGTACCCCATMGAGACTCTCAGAG 1549  
 QY 1872 accctctgagcaactgagatccttccagagtcaggagatccaggttgaatacaaga 1931  
 Db 1550 ACCCTCTGAGCAACTGAGCTATCTTCCAGAGTCCAGGAGTTCAGGTGATTAACAAGA 1609  
 QY 1932 tttccccaaaaaacaagaagaagattgtatctcttatacaatbgtcctccagccacc 1991  
 Db 1610 CTTCCCAAAAACAACAAGAAAGAAATTTGTATCTCTTATCAATCTCTCTCAGCACCC 1669  
 QY 1992 tctgataagccaatglatcgcaagagatgtagtctctgccaatgatatggtcgtgctga 2051  
 Db 1670 TCTGATCAGCCATGTTATGCGCAAGAGATGTGAGTCTCTGCCATGATATGGCTGCTGAA 1729

QY 1451 tgcctgagatcccttgccttcaaaagctcccgagcggcagcccaaccacaacccgcactcaagt 1510  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1087 TCCTGGAGATCCTTGTTCAAGTCCCGCAGCGGACGCCACCAAAACCCGCACTCAAGT 1146  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 1511 cagagagagaagaaacccaataaagaacccggggaatggaagaagaatgaactcttttgaac 1570  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1147 CAGAGAGAGAAAGACCCATTAAGAAACCAAGGATGTAAGAAAGTAACCTTTTGTGAAAC 1206  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 1571 ctggcctcggggaatgaataatggaactgaataaagaagaatgaatcagaatgcctatc 1630  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1207 CTGGCTCTGGGGATGAAATGGACCTAGTAATAAGAGATGATGATTCAGATCCCTTATC 1266  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 1631 taagtcataagcaagctgcctcctggaattctcccaatgctgcggaagctgcgcaagctg 1690  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1267 TAAGTATAGCAGCTGCTGCTGGAATTCCTCCATGCGCCGAGAGTCCGCAAGCTG 1326  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 1691 taggaatgtaagaagaatcaacaaagaatttaaccaggagctccgaatccctgcga 1750  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1327 TAGGAGTATAGTCAAGGACATCACCAAGATTTTACCAAGGACGCTCCGAATCTGCCA 1386  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 1751 aagccacgctaacctgcataatagccggaagtgtgtataggggcaactgcgccaag 1810  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1387 AGCCCAAGGTAAGTGGCATGATAGCCGAGAGTGTGTATGGGGCACCTCGCCACAG 1446  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 1811 ccgaagacatlttaagaataacatctctcaaggcaagctaccccaatggaactctcaaga 1870  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1447 CCGAGACCAATTTTAAGAAATACATCTTTCAGGCCACGTAACCCCATGACCTCTCAGCA 1506  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 1871 gacccctcgaagcaactggaactatcttccagaatccagaagatccgaagtctgaataaag 1930  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1507 GACCCCTGAGCAAGTGAAGTATCTTTCAGAGATTCAGAGGATTCAGAGTGAATACAAAG 1566  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 1931 acttccccaaaaacaagaagaaatctgtaactcaatctcctcctcaagcaac 1990  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1567 ACTTCCCAAAAAACACAGAGAGAGATTTGTATCTTATCAATTCCTCTCAGCCAC 1626  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 1991 cctcgtacgaatggtatctcgaagaatgctggaatctcctgcataatgctgcgctga 2050  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1627 CTTCTGATCAGCCATGATTCGCAAGAGATGTGAGTCTCCATGATATGAGTGCCTGA 1686  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 2051 acatcttaagtctgctgtcgtgagcttgaccacaagaatcaagaatgccaagaacggaa 2110  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1687 ACATCTTAAAGTGTGCTGTGATGAGTGGACCAAAAGTACAGAGATGCCAAGAACGGAA 1746  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 2111 acggaacaaatcctgctgtctggaaggtgctgaaactcttctgcgaatgaacatlaaa 2170  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1747 ACGGACCAATGCTGTGTGTGGAGAGTGTCTGAACCTTTTCTGGCCATGAACCTTTATAA 1806  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 2171 atcccaacataatactgaanaatactgaactgcttgaanaatttgaattctgaatc 2230  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1807 ATCCCAACATATATCTGAATAATGTAACCTGCTTTGAAATTTGGAATTTCTGATACC 1866  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 2231 tccgaatgggcccgaagaacacggttgggttaagaatgttgggcaagcaaggaagaacaacag 2290  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1867 TCCAGTGGGCCGAGACACGCTGTGTAAGATGTGGCAGCAGAGAGGAACAAACAG 1926  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 2291 aaacacaagaagcggctctgagccggctggaactgctggaagtcttgggtcttggatggcaact 2350  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1927 AAACACAAGAGGCGGCTGTGGCCGCTGACTGTGCTGGGGTTTGTGTGATGGCCACT 1986  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 2351 cggctgaacctggcggctcctcactgaacataagaacgtgcctgttgggaagaagggctgcagc 2410  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1987 CGGTGACCTGGCGGTCCCTACGCAATAGCAGTGCCTGTGGGGAAGAAGGCGTGGCCAGC 2046  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 2411 cagctgttctcccggaacacagcaatccaaacccctgggcaacctccgtgttgcctc 2470  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 2047 CAGCTGTTCTCCCGGAGACACGAGATTCACACCTGGGACCTCCGTTTGGCTT 2106  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 2471 tttttccctctgttgaagaagaacggcagaaccccttcaaaactggctcaactcaga 2530  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 2107 TTTTTCCTCCCTGTGTGAAGAGAGAAACGGACACACCCCTTCTAAGCTGCTACTCAGA 2166  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 2531 cacatctggagacaacacctgacagcattgcagaagagagccttgcacggccccagagc 2590  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 2167 CACATTGGGACAAACCTTGACAGCCATGCGAGAGAGAGCCTTTGACCGGCCACAGC 2226  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 2591 taaagcaccagaagaataatgaatgcttccactcagcggagccaaacttctagtg 2650  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 2227 TAAAGACCGAGAAATATAAATGCTTCTACTACAGCTGAGCCACTTTTCTAGTG 2286  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 2651 ccagggcccccacacactcctcgcagtaaccacacatcaacacactgcttctctccaag 2710  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 2287 CCAGGGCCACACACCTCCGTCAGATGCCACACATCACCACAGTCTTCTTCCAAACAG 2246  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 2711 tgaatcgtatctcctagtttcaataattcttcttgatgataatgaactataaattc 2770  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 2347 TGATCTGATATCTTAGTTCATATATTTCTTTTGTGATGATGACACTATATATAATTT 2406  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 2771 catttgaagaattctcaatgtaactgaatgaataaagaacagcttgaagaacttctgaga 2830  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 2407 CATTGGAATTTCTCAATTTGATCTATGTTAAATAGCACAGTTTGGAACTTGTCTGAGA 2466  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 2831 ctgacttatacaataactaaccgacaagaatcalatcaatgtaatgcttgaacat 2890  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 2467 CTGACTTTATCATATATCTAATACCGACAAAGATCATATCATATGATGAGTATAGCAT 2526  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 2891 tttatctcaatgtaacccaaggaacagttcagtgatgaataatgctgctccctggtc 2950  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 2527 TTTATTTTCATTTGACTTAACCCAGAGACAGTTTCAGTGAATGCAAAATGTGTGCTGCTC 2586  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 2951 agctgaagaagctccttggaacttcaaaaacttgaataagctcctcgaagctgtataat 3010  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 2587 AGCTGAACAGTCTGTGACTTTCAAAAACCTTGAATAGTCTCCACAGTTGTATTAAT 2646  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 3011 ggaacatttgaagaatttlaaacttgaatgatcatcttgcatttcaatttcaattc 3070  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 2647 GGACAAATTTTGAATTTTAACTTTAGATGATCATTTTGTTCATTTTATTTTCAATTTT 2706  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 3071 atttctgttaatgaacaacaggaacttgaatgaacttgcattcgtttaaagaattata 3130  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 2707 ATTTTGTATATCAAAACAGGACTTAATGAACTTTGATCTCTGTTTAAAGATTATTA 2766  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 3131 aaaaacttggtaactatataatagctccttgaaggacttgaacttcaactcaactaga 3190  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 2767 AAAACATTTGATCTATATATATATGCTCTTGAAGACTTTCATCTACACTACAGGA 2826  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 3191 tatgatctcagtgtaagccataataaactgcagagtgatcttccagaatgctcgaactg 3250  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 2827 TATGATCTCATGTAGTCCATATTAACCTGCAGAGATTTTCCAGAGTCTCGATCTG 2886  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 3251 ttaataacatctccatlaaggctggaagaagaatgaactgaatgcttctgatacagctgtc 3310  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 2887 TTAATTAACATCTCCATTAAGGCGTAAGAAAGTAAGACCTTCTGTATACACTGTGTT 2946  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 3311 gcttctgaatgcttgaatcgtatcagtaacagaagtgctgcaactggaagctcgtgtgtcc 3370  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 2947 GCTTTGATGTTGTGTACTGTACACGAAAGTGTGTGACTGAGGCTCTGCTGTGTCTC 3006  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 3371 gtaagaaaacctcgtagccctcgcagtgtaagtaactgctcactcaatgcttctgaactg 3430  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 3007 GTATGGAACCAACGTGATAGCCCTCGAGTTAAGTATGCTTCCATTCATGTTTAAACGCTG 3066  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 3431 aatttctcccaatggaatgaatgaataaactcaagtgttctgcatcaataatgtaact 3490  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 3067 AATTTTCTCCCATGGAATGTAAGTAAACCTTAAGTGTGTGATCAATAAATGTGAAT 3126  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 3491 acttaaaaaaaaaa 3506  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 3127 ACTTAATAAAAAAAAA 3142  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 5  
 AK025519 AK025519 3194 bp mRNA PRI 29-Sep-2000  
 LOCUS Homo sapiens cDNA: FLJ21866 f1s, clone HBP02379, highly similar to  
 DEFINITION AF061938 Homo sapiens stauufen protein (STAU) mRNA.

QY	3061	tttatatttatatttggtaatgcaaaaggaactaaatgaacttgatctcgtgtttaa	3120
Db	2772	ttttcatTTTTATTTTGGTTAAATGCAAAAGGACTTAAATGAACTTTGATCTCTGTTTTAA	2831
QY	3121	agatlatcaaaaacactgtgtatcatatacatatgagctcttgagagacttgaacttca	3180
Db	2832	AGATTATTTAAAAACATGTGGTATGATCATCATATAGGTGCTTGGAGACTTACCTTCACTA	2891
QY	3181	caactacaggaatgatgatctccatgtagtgcataataaacctgcagagtgatlttccagagt	3240
Db	2892	CACATACAGGATTAAGACTCTCCATGTATGATCCATTTAAACCTGGAGAGGTATTTTCCAGAGTG	2951
QY	3241	ctcgatactgttaattacatctccatcttaagagcttgaaagaatgtgacctacgtcttcgtata	3300
Db	2952	CTCGATATCTGTTAATTATACATCTCCATTAAGGCGTGAAGAAAGTAAGCATACGTTTCTGTATA	3011
QY	3301	caagctgttgttgcctttgatgatgtgtgttactgtacacagaagaatgtgtgactgaagctctg	3360
Db	3012	CAGCTGTGTTGCTTTTGAAGATGTTGTATTACTGTACACAAAGAGTGCTGACACGAAGCGCTG	3071
QY	3361	cggtgtgtctcgatctggaaaaacccgtgtgaagccctgcgaagtttaagtactgcttccattcgtg	3420
Db	3072	CGTGTGATCCGTAATGGAANAACCTGGTAGGCCCTGCAAGTTAAGTACTGTTCCATTTCATG	3131
QY	3421	tttaacgttggaatttctctcccatctgaaatgtgaatgaaacttaagtttctgcatcaat	3480
Db	3132	TTTACGCTGGAAATTTTTCCTCCCATGGAATGTAAAGTAAAGCTTAAGTGTGTCATCAAT	3191
QY	3481	aaatggttaataactcaaaaaaaaaaaaaa	3506
Db	3192	AAATGTTAATATCTPAAAAAATAAAAA	3217

RESULT	4
LOCUS	AF061940
DEFINITION	AF061940 3142 bp mRNA PRI 04-MAR-1999
ACCESSION	Homo sapiens staufen protein (STAU) mRNA, alternatively spliced, complete cds.
VERSION	AF061940
KEYWORDS	AF061940.1 GI:4335950
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 3142)
TITLE	Wickham, L., Duchaine, T., Luo, M., Nabl, I. R. and Desgroselliers, L. Mammalian staufen is a double-stranded RNA and tubulin binding protein which localizes to the rough endoplasmic reticulum
JOURNAL	Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
MEDLINE	99147057
REFERENCE	2 (bases 1 to 3142)
AUTHORS	Desgroselliers, L., Wickham, L. and Luo, M.
TITLE	Direct Submission
JOURNAL	Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900
FEATURES	Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada
source	location/Qualifiers
	1..3142
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="20"
	/map="20q13.1"
	1..3142
	/gene="STAU"
	288..1778
	/gene="STAU"
	/function="RNA binding protein"
	/note="alternatively spliced transcript"
	/codon_start=1
	/product="staufen protein"
	/protein_id="AA017532.1"
	/db_xref="GI:4335951"

```

/translation="MKLGKKPMYKVPDYSRMSTNYNNRGGAYPPRYTYPPPPPL
LYVLSVGGQOENGKTRQAAKHDAANALILQEPYPERLYVGRSSEENLAK
SELSOYVEIALKRLPVNEFVARESGPPHKNKXITQVGEFVGGGKSKISKNA
AIVLEELKTLPLPAVERVKPIKKTKEITVPQSPEDGQINPLSLAQIOOAK
EKBEYTYLTERGLPRRREFVMOYKCNHIAECTGNKYAKRANENMELITGFVP
QRODTPALKSEETPIKPKGDGKRKYTFEPGSGDENGYSNKDEPRMPLSQO,LA
GLIPMPDEVAAQVSGGHTKDTFRAAPNAPAYAMALVGGSTPYMETLIL
NNISGHPHGPLTRPESQDLYSRQGFVFAEKDKPKNNKNEFVSLINCSOPPLS
HGIGKDVESCHDMAALINILKLSLDDQSTEMPRTQNGRPSVCGRC"

```

Query Match	82.6%;	Score 2896;	DB 10;	Length 3142;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2896;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	611	caaaagcatcaacccctactgtgtgaacttaatgaactgtgcatgtgaactgtgaaaaaac	670
Db	247	cagaaacatataccctactctgtatgaacttaatgaactgtgcatgtgaactgtgaaaaaac	306
Qy	671	caatgtaataagcctgttgacccttactctcgtatgagtcacccataactacaatgta	730
Db	307	caatgtataaagccctgttgacccttactctcgtatgagtcacccataactacaatgta	366
Qy	731	gaggaagtgcttatcccccgaggtactttaccatttcagtttccaccttacttatac	790
Db	367	gaggaagtgcttatcccccgaggtactttaccatttcagtttccaccttacttatac	426
Qy	791	aagtggaactctctgtgtggagagcaacatlaatgagcaagaagaacaagacagctg	850
Db	427	aagtggaactctctgtgtggagagcaacatlaatgagcaagaagaacaagacagctg	486
Qy	851	cgaacaacagatgctgtcgtccaaagcgttgagatccctcgaaatgagccctgcagaga	910
Db	487	cgaaacacagatgctgtcgtccaaagcgttgagatccctcgaaatgagccctgcagaga	546
Qy	911	ggctcgagagtggaatlgtaagaagaaatccgaagaagaataatcgaataactctgaataagtc	970
Db	547	ggctcgagagtggaatlgtaagaagaaatccgaagaagaataatcgaataactctgaataagtc	606
Qy	971	aagtgcttgagatgtcacttaaaacggaactgtccttgaaatttcgaagtggcccgagaga	1030
Db	607	aagtgcttgagatgtcacttaaaacggaactgtccttgaaatttcgaagtggcccgagaga	666
Qy	1031	gtggccaccccccaatgtaagaacttgtgaccaaagtltcogtltggggagtlgtgtggg	1090
Db	667	gtggccaccccccaatgtaagaacttgtgaccaaagtltcogtltggggagtlgtgtggg	726
Qy	1091	aagtgtaagggaaaaagtaagaatltcaagaagaaatgctgcgtatagctgtlcttgagg	1150
Db	727	aagtgtaagggaaaaagtaagaatltcaagaagaaatgctgcgtatagctgtlcttgagg	786
Qy	1151	agctggaagaagttacccgcctcgtcgtcagttgaaagatlaaagcctatgaatcacaaga	1210
Db	787	agctggaagaagttacccgcctcgtcgtcagttgaaagcctatgaatcacaaga	846
Qy	1211	aaacaaaacccatagtcgaagcacaagacaagcccgaaatlatgcccagggatcaatccga	1270
Db	847	aaacaaaacccatagtcgaagcacaagacaagcccgaaatlatgcccagggatcaatccga	906
Qy	1271	ttaagcagacttgcccgatcccaagcagagcaaaaaaggaaggaagagccagagatcaacgctcc	1330
Db	907	ttaagcagacttgcccgatcccaagcagagcaaaaaaggaaggaagagccagagatcaacgctcc	966
Qy	1331	tcacagagcgagagcctcccgccgcgcgaagggagtlttgtga tgcaggtgaaagtltgnaaac	1390
Db	967	tcacagagcgagagcctcccgccgcgcgaagggagtlttgtga tgcaggtgaaagtltgnaaac	1026
Qy	1391	acaactgcaagaagaacgagagcaacaagaagaagtggtgccaagcgacatgcaagcagagaaa	1450
Db	1027	acaactgcaagaagaacgagagcaacaagaagaagaagtggtgccaagcgacatgcaagcagagaaa	1086

```

Db 552 AGACAGGCTCGAAACGATGCTGCTGCCAAAGCGTTGGAGATCCGACGAATGACCCC 611
Qy 901 ctcccaagagcgctgagagtgaaatgaaagaatccgaagaagaataatccaataatc 960
Db 612 CTGCCAGAGAGGCTGGAGGTGAATGGAAGAGATCCGAAGAAGAAATCTCAATAATCT 671
Qy 961 gaaataagtaagtgatttgagattgcacttaacggaacttgcctgtgaatttcgaagtg 1020
Db 672 GAATAAGTCAAGTGTGTGAGATTGCACCTTAACGGAACTTGCCCTGTGAATTTGAGGTG 731
Qy 1021 gcccgagagagtgcccaaccccaatgaagaactttgtgaccaaagtttcogtgggag 1080
Db 732 GCCCGGAGAGTGGCCACCCACATGMAAGACTTTGTGACCAAGTTTGGTTGGGGAG 791
Qy 1081 ttgttgagggaagtggaaggaagaattccaagaaataatgcccataagct 1140
Db 792 TTTGTGGGGAGAGGTGAAGGGAAGAAAGCAAGATTTCAAGAAATGCGCCATAGCT 851
Qy 1141 gtctctgagagctgaaagaattaccgccccctgtcctgcaagtgaacgaataaagcctaga 1200
Db 852 GTTCTTGAGAGGCTGAAGAAGTTACCGCCCTGCTGCAGTTGAACGAGTAAGCCTAGA 911
Qy 1201 atcaaaaaaagaacaaacccatagtcagccacagaacagcccaagataatgycgaaggg 1260
Db 912 ATCAAAAAGAAAAACAAACCATTAGTCMAAGCCACAGAACGCCAGAAATATGGCCAGGGG 971
Qy 1261 atcaatccgattagccagctgcccagatccagcagcgaacaaagagagagccagag 1320
Db 972 ATCAATCCGATTTAGCCACATGGCCCAATCCAGAGGCAAAAAAGGAAGACCCAGAG 1031
Qy 1321 tacaagctcccaacagaagcagagcctcccgccgcagcagagagtttgatgacgagtaga 1380
Db 1032 TACACGCTCTCAACAGAGCGAGGCTCCGCGCGCAGGGAGTTGTGATGCAAGTGAAG 1091
Qy 1381 gttggaacacacactgcagaaggaacgggcacaaagaagtgycgaagcgaatgca 1440
Db 1092 GTTGGAAACCCACTGCAGAGGAAGAACGGGCACCAAGAAAGTGCGCCAAAGCCGAATGCA 1151
Qy 1441 gccgagaacatgctgagatccttggtttcaaaagtcgcgagcgagcccaacaaacc 1500
Db 1152 GCCGAGAACATGCTGGAGATCTTGGTTCAAAAGTCCCGACCGGCCACCAAAACC 1211
Qy 1501 gcactcaagtcagagaggaagacacccataaagaaccaaagggatgaaagaaatlaac 1560
Db 1212 GCATCTCAAGTCAGAGGAGAGACCCATTAAAGAACCAAGGGATGGAAGAAAGTAACC 1271
Qy 1561 ttttttgaacctggtctctgggagatgaaatggaatagtaataaagaagatgagttcag 1620
Db 1272 TTTTGTGAACCTGGCTCTGGGGATGAAATAAGGACTAGTAATAAGAGATGAGTTGAGG 1331
Qy 1621 atgacctatataagtaacgagctgctgctggaatttcccaatgctgcccagagtc 1680
Db 1332 ATCCCTTATCTAAGTCATCGACAGCTGCTGCTGGAATTTCTCCATGAGTGCAGAGGTC 1391
Qy 1681 gcccaagcgtgtagagtgtaatcaagaacatcacacaaagaatttaccagggcagctccg 1740
Db 1392 GCCCAGGCTGTAGAGATTACTCAAGGACATCACCAAAATTTTACCAAGGCAAGCTCCG 1451
Qy 1741 aatctgccaaggaacagtgtaactgcatagtagccgagagtggtgtatggyggacac 1800
Db 1452 AATCTGTGCCAAGGCCACAGGTACTGTCATGATGCCGAAGATGTTGTATGGGGGCACC 1511
Qy 1801 tgcgccaagcgagagacctttaaagaataaatctcttcagggccacgtaacccatgga 1860
Db 1512 TCGCCCAACACCGACCAATTTTAAAGAAATAACATCTCTTCAGGCCCGATCCCATGGA 1571
Qy 1861 ccctcaacagacccctctgccaactgactatcttcagagatccagggatcccaagtt 1920
Db 1572 CCTCTCAGAGACCTCTGTAGCAACTGATATCTTTCCAGAGTCCAGGAGATTCCAGGTT 1631
Qy 1921 gaataaagaacttcccaaaaaacaagaagaaatgtglatctctatcaatltgctcc 1980
|||||
Db 1632 GAATACAAAGACTTCCCAAAAACACAGACGAATTTGTATCTCTTAATCAATTGCTCC 1651
Qy 1981 tctcagcacctctgatcaagccatglatcggcaaggaatgtagagtcctgtccatgatacg 2040
Db 1692 TCTCAGCACCTCTGATCAGCATGATCGCAAGAGATGTGAGTCTCTGCATGATATG 1751
Qy 2041 gctgcgtgaaacatcttaaaagtgcgtctgaglttgaaacaaagaatcacagagatgcca 2100
Db 1752 GCTGCGCTGAACATCTTAAAGTTGCTGTGAGTGTGAGACCAAAAGTACAGAGATGCCA 1811
Qy 2101 agaaagaagaacgagcaacaaatgctgtgtgtggaggggagtcgtaaccttttcggcaatgaa 2160
Db 1812 AGAAGAGAAACGGACCAATGCTGTGTGTGGAGGTGCTGAACCTTTTGTGGCCATGAA 1871
Qy 2161 ccaatlaaataatcccaacataataactgaaataactgaactgtcttgaaatltggaat 2220
Db 1872 CCATTTAATAATCCCAACATATATACGMAAATCTGAACCTGTTTGAAATTTGGAGAT 1931
Qy 2221 tctctaacctccagtgaggcagagacaagtgagtaaaagatgtyggcagcagcaggg 2280
Db 1932 TTTGTATACCTCCAGTGCGCCGAGAGACAGGTGGTAAAGATGTGGCAGCAGAGG 1991
Qy 2281 aagaacaacagaacaaagaagcgctgtgtgcccgtgtgacgtgtgctgtggtgtgtgtc 2340
Db 1992 AAGCAACACAGAAACACAAGAGGCGGTGTGGCCGCTGGACTGTGCTGGGGTTTGTGT 2051
Qy 2341 gatgacacctcgtgtgacctgagcgtgtccctacgcaatagcaactgtcgtgtggagaagag 2400
Db 2052 GATGGCCACTCGGTGACTGTGGCTGCTTACGCATATAGCACTCTCTGTGGGAGAAAGG 2111
Qy 2401 gctgccccagcagctgtgtctcccgagacacagacagatccacaccccttggycacctcgt 2460
Db 2112 GCTGCCAGCAGCTGTGTTGCCGGGACACACAGATGCACACCTGGGCACTCCGCT 2171
Qy 2461 gtttggtcttttttccctgtgtgaaagaagaacggcagcagccctctcaagctgtg 2520
Db 2172 GTTGTGCTTTTGTTCCTCGTGTGAAGAAAGAGCGCACGCCCTTCTCAAGCTGG 2231
Qy 2521 ctcactcagacacatlggagcaaacaccttgagacagccatgccaagagagagccttgagcg 2580
Db 2232 CTCACCTAGACACTTTGGGACAAACCTTGACAGCCATGCCAGAGAGAGGCTTTGACCG 2291
Qy 2581 gccccaagactlaaaagcagcagagaagaatacaatgcttccactcaagctgacccaact 2640
Db 2292 GCCCCAGAGCTAAAAAGCACAGAGAAATCAAAATGCTTCTACTACAGGTGACCAACTT 2351
Qy 2641 tctcagtgctgccaagggcccaacacacccctcctgcaataccacacataccaactgtcttc 2700
Db 2352 TTTCTAGTGTGCCAGGCGCCACCACTCTGCACTACCCACACATACCACTGCTTTCT 2411
Qy 2701 ctccaacagtgatcgttatctctagtttcaataatcttcttcttgatgataagcactat 2760
Db 2412 CTTCCAACAGTGAATCTTATTTAGTTTCATTTATTTTCTTTTATGATATGACACTAT 2471
Qy 2761 aataaatlttcaatltgaaatlttccaatltglatcgtatgaaatlaagacaglttgaaac 2820
Db 2472 ATAAATTTTCATTTGGAATTTCTCAATTTGTATCTGATTAATATAGACAACTTTGGAAC 2531
Qy 2821 tctgtcgaagactgacttatacaataactcaacgcaagaagataatccatgtgtatgtg 2880
Db 2532 TTTGCTGAGACTGACTTTATCAATAAATCTTAACCAAAAGATCTTAATTCATGTCTAATG 2591
Qy 2881 gttagacatttataatcatctgactgaacccagagacagtttcaagtgtgacaatgtgtgc 2940
Db 2592 GTTAGACATTTTATTTTCAATTGACTTAACCCAGACAACTTTTCAGTGAATGTTGTGTC 2651
Qy 2941 ccctctgctcagctgaaacagctcctggaacttcaaaaactltgaataagtcctccacagt 3000
Db 2652 CCTGTGTTTACGCGAAGAACGTCCTGAGCTTTCAAAAAACCTTGAAATTAAGTCTCCACAGT 2711
Qy 3001 tgtataaatltggacaactltggaatlttaactltgatagtatcattgtgttcaatlttta 3060
|||||
Db 2712 TGTATAAATTTGACAAATTTAGAAATTTAAACTTTAGATGATCATTTTGTTCATTTTTA 2771
```

QY 3061 ttcaatttatttctgttaacgaacaggaacttaagtaacttgatctcctgtttaa 3120  
|||||  
Db 2904 TTTCAATTTTATTTTGTGTAATGCAACAGACGTTAAATGAACTTGATCTGTGTTTAA 2963  
QY 3121 agatataaaaaaacatctgtatctatcatatagctctcttgaggaacttaagcttca 3180  
|||||  
Db 2964 AGATATATTAACCAATGTTGATCTATACATATGCTCTTGAGACCTTAGCTTACACTA 3023  
QY 3181 cactcaagaatataagctcccatctgtatgctcatataaactcgagagtgatcttccaga 3240  
Db 3024 CACTACAGAAATATGATGATCCATGATGTCATATATAAACCCTGACAGATGATTTTCCAGAGTG 3083  
QY 3241 ctgcatactgttaatacatctccatctagagctgaaagaagaactgaactcgtttcgtata 3300  
Db 3084 CTCGATCTAGTTAATTAATACATCTCCCTTAGGCGCTGAAGAAATGACCTTACGTTTCTGTATA 3143  
QY 3301 cagctgtgtcttcttgatctgtgtgttaactgtacacagaagtggtgtaactgaagctctg 3360  
Db 3144 CAGCTGTGTGCTTTTGTGATGTTGTGTACTGTACACAGAAAGTGTGACACTGAGGCTCTG 3203  
QY 3361 cgtgtgtccgtatgaaacacagctgtagccctcgagagtgatgactgttctcatctatg 3420  
Db 3204 CGTGTGTCCGATGAGAAACCTGATGACCCCTGCGAGTTAATGCTGCTCCATTTCATTG 3263  
QY 3421 tttaagctggaatttcttcccatggaatgtaagtaaaacttaagtggttgcatacat 3480  
Db 3264 TTTACGCTGGAAATTTTCTCCCATGGAATGTAACTAACTAAGTGTGTTCTCATCAAT 3323  
QY 3481 aaatgttaactactaaaaaataaa 3506  
Db 3324 AAATGTATACTATAAAAAA 3349

RESULT 3  
AF061938 3217 bp mRNA PRI 04-MAR-1999  
LOCUS AF061938 Homo sapiens stauflen protein (STAU) mRNA, alternatively spliced,  
DEFINITION complete cds.  
ACCESSION AF061938  
VERSION AF061938.1 GI:4335946  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 3217)  
AUTHORS Wickham,L., Duchaine,T., Luo,M., Nabi,I.R. and Desgroselliers,L.  
TITLE Mammalian stauflen is a double-stranded RNA and tubulin binding  
protein which localizes to the rough endoplasmic reticulum  
JOURNAL Mol. Cell. Biol. 19 (3), 2220-2230 (1999)  
MEDLINE 99147057  
REFERENCE 2 (bases 1 to 3217)  
AUTHORS Desgroselliers,L., Wickham,L. and Luo,M.  
TITLE Direct Submission  
JOURNAL Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900  
Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada  
FEATURES  
SOURCE Location/Qualifiers  
1..3217  
/organism="Homo sapiens"  
/db\_xref="Laxon:9606"  
/chromosome="20"  
/map="20q13.1"  
1..3217  
/gene="STAU"  
/gene="STAU"  
363..1853  
/gene="STAU"  
/function="RNA binding protein"  
/note="alternatively spliced transcript"  
/codon\_start=1  
/product="stauflen protein"  
/protein\_id="AADI7530.1"  
/db\_xref="GI:4335947"  
/translation="MKIGKKPMYKVPDPYSNMOSTYNNMNGAVAPRYEYFPVPLP

LYOVELSVGGQOFNGKGTROAKHDAAKALRIIONEPLPERLEVNGRESEENLNK  
SEISOVEILKRLNLPVNFVARESPGPHKKNFPTKVSFEPFVEGSGSKISKKA  
AIALVEELKTRPELPAYERYKPRIRKKTTRVYKQTSPEVGGGINSPLRAIQDAK  
EKEPEYTLTERGLPRRERFVMOVKVNHTAREGVTNKKVAKRNAENMELISFKVP  
QROPTKPAKSEKPTIKPDGKRKVFEPGSGDEGTSNKEDEFPMLPSHQDLA  
GILPMEVADAVSOGHNTKDPTRAPNPNAKTAVAMLARLELYGCTPTEETIK  
NMISGHPHPLTRPSEOLDYLSRQDFOVEYEDFPKNKNNEVSLINCSQPLIS  
HGICKDYESCHDMALNLIKLSRLDQOSTEMPRTGNPSVGCRC"  
BASE COUNT 889 a 796 c 742 g 790 t  
ORIGIN  
Query Match 83.1%; Score 2914.8; DB 10; Length 3217;  
Best Local Similarity 91.7%; Pred. No. 0;  
Matches 3215; Conservative 0; Mismatches 2; Indels 289; Gaps 1;  
QY 1 acttctgcgggctgtcgggcgctgagcgctctcaagcglttgcgggcggtgcgt 60  
|||||  
Db 1 ACTTCTGCCGGGCTGCGGGCGCTGAGCGCTTCAACGGTTTGGCGGCGCTCGCGT 60  
QY 61 ctctctggctcccgcttctctttagacgcgtcccccgcggcgccggcgccgcctc 120  
|||||  
Db 61 CTCTCTGGCTCCGCTTCTTGAACGCTCTCCGCCCGGCGGCGGCGGCGGCGCTC 120  
QY 121 ctccagggcaactccgcctcttccctctctctctctctctctctctctctct 180  
|||||  
Db 121 CTCACGGGCACTCGGCTCTTCCCTCTGTCCTCTTCTCTCTCTCTCTCTCTT 180  
QY 181 ctctctccctccctccgcgcgcacccgcgcgcgcgcgcgcgcgcgcgcgcgcgc 240  
|||||  
Db 181 CTTCCTCCCTCTCTGCGCGCGCACCGCCAGGACCGCGCGCGGCGGAGAGTCCGAGC 240  
QY 241 agcaagccagagtttattcaaccacttaacctctcaagactgaaacaaagcaatgttcc 300  
|||||  
Db 241 AGCAGCCAGAGTTTATTAACCACTTAACCTTCAGAACTGAAACAACAACTTTGTTCC 300  
QY 301 tggagcgccctcttcttaaaaaagtagaacttagacttcaagactgaactaactg 360  
|||||  
Db 301 TGGAGCGCCCTCTTTTAAAAA----- 322  
QY 361 cactgaaagctgttlaacctgttaactgttcaacttctgtgaaagtagaactgttcaagtc 420  
|||||  
Db 323 ----- 322  
QY 421 caagtgcaggttcaagaaccactctgtctctctcgaaggagccaatactgacaagaagc 480  
|||||  
Db 323 ----- 322  
QY 481 cagtccttctctcaagccttgatgagtattccttctactactagctctcgcctct 540  
|||||  
Db 323 ----- 322  
QY 541 gaaatgaagtagaaccatccaactcaactctgttaacctgtcatctataatccaacc 600  
|||||  
Db 323 ----- 322  
QY 601 agtgcagctgcagaagaataaccctactctagaaactaaatgcactgtgcatgaact 660  
|||||  
Db 323 -----AGAAAGCATACCCCTACTGTAACTTAAAGCACTGTGTAAGACTT 371  
QY 661 ggaataaacaactgataagcctgtgaacctactcctcgaatcagctccaccataac 720  
|||||  
Db 372 GGAATAAACCAATATATAGCCTGTGACCTTACTCTCGATGCAATGCCACTATAC 431  
QY 721 tacaacatgagagagtgctatcccccgagtgacttcaaccattccagttccacct 780  
|||||  
Db 432 TACACATGAGAGGAGGAGGCTATATCCCGAGAGTACTTTTACCATTCACATTCACACT 491  
QY 781 ttactttatcaagtggaactctgtgaggagagcaacttaatgagcaagaagaaga 840  
|||||  
Db 492 TTACTTTATCAAGTGGAACCTTCTGTGGAGAGCAGCAATTTAAAGGCAAGAAAGACA 551  
QY 841 agacagcgctgcgaacaacagatgctgtgcacaagcgttgagagatccttgcaaatgagcc 900



Db 684 AGACAGCGTCGGAACACGATGCTGCTGCCAAGCGTTGAGATTCCTGCAGAAATGAGCCC 743  
Qy ctcgcaagaagcctggaagtgaaatggaagaatccgaagaagaataatccaaataatc 960  
Db 744 CTCGACGAGAGCGTGGAGTGTAATGGAAGATATCCGAAACAGAAATATCATTAATCT 803  
Qy 961 gaataaagtcgaatgtttggaatgcactaaacggaacttcgctgtaatttcgaagt 1020  
Db 804 GAATATAGTCAATGTTGATTTGACTTAACGGAACCTTGCTGTGATTCGAGGTG 863  
Qy 1021 gcccggaagaagtcgcaacccacataagaacttctggaagaagtcttggttggaag 1080  
Db 864 GCCCGGAGAGTGCCACCCACATGAAGACTTTGTGACCAAGTTTGCTGGTGGGAG 923  
Qy 1081 ttctgtggaagaagtcgaagaagaagaatctcaagaagaataatgcgcacatagct 1140  
Db 924 TTTGTGGGGGAAGTGAAGGAAAGCAAGACATTTCAAGAAATAATGCCGCATATGCT 983  
Qy 1141 gtctctggaagagctggaagaagttaacgcgcctgctgcaagtgaacgaataagcctaga 1200  
Db 984 GTTCTTGAGAGCTGAAGAAGTTACGCCCCCTGCTGCAGTTGACGAGTAAGAACCTAGA 1043  
Qy 1201 atcaaaaagaacaaacacacatagttcaagccacagaacaaagcccaagataatggtccaggg 1260  
Db 1044 ATCAAAAAGAAAACAAAACCCATATGTCAAAGCACAGCAAGCCACAGATATGCGCCAGGG 1103  
Qy 1261 atcaatccgaattagccagactggtccagatccagcagcaagaagaagaagagccagag 1320  
Db 1104 ATCAATTCGATTTAGCCGACACTGCGCCAGATCCACAGGCAAAAAAGGAAGAGCCAGAG 1163  
Qy 1321 taacagctccacagagcgaagcctccgcgcgcgaaggaagtctgtatgcagtgtaag 1380  
Db 1164 TACACGCTCTCTACAGAGCGAGCGCTCCGCGCGAGGAGTTTGTGATGACGAGTCAAG 1223  
Qy 1381 gtctggaacacacacctgcagaagaacggtgcacaaagaagaatggtccaaagcgaatgca 1440  
Db 1224 GTTGGAAACCCACTGTCAGAAAGAACGGGACACCAAGAGTGGCCAAAGCCCAATGCA 1283  
Qy 1441 gccgaagaacatgcctggaatccttggcttcaagaatcccgagcgaagcccaacaaacc 1500  
Db 1284 GCCGAGAACATGCTGGAGATCCTTGGTTTCAAGTCCCGACGCGGACGCCACCAAAACC 1343  
Qy 1501 gcaatcaagtctcagaagaagaacacacataagaagaacagggagtcgaagaagaatcaac 1560  
Db 1344 GCACCTCAAGTCAGAGGAGAAAGACACCCATTAAGAAACAGGAGTGAAGAAAGTAACC 1403  
Qy 1561 tcttttgaacctggtccttggaatgaaatgggaactagtaataagaagaatgagttcgaag 1620  
Db 1404 TTTTGTGAACCTGCTCTGGGATGAAATGGGACTAGTAATAAAGAGATGAGTTGACAG 1463  
Qy 1621 atgacctatcattatcagaagcgtgcctggaatcttcccatgtgtgccgaagtc 1680  
Db 1464 ATGCTTATCTTAAGTATCAGCAGCTGCGCTGTGAATTTCTCCATGGGCGCCGAGGTC 1523  
Qy 1681 gcccaagcctgtaagatcgaatcagaacatcacacaagaatcttcaagggcagactcgc 1740  
Db 1524 GCCAGGCTGTAGAGATTAAGTCAAGGACATCACACCAAAATTTTAACAGGAGAGCTCCG 1583  
Qy 1741 aatctgcgaagcgaacggttaactgcacatgataagccgaagatctgtatgtaggggaccc 1800  
Db 1584 AATCCGTGCCAAGGCCACGGTAACTGCGATGATGAGCCGAGATTTGTTGATGGGGGACCC 1643  
Qy 1801 tgcgcgaagcgaagaacctttaaagaataacatctctcagaagcaagtaaccccatgga 1860  
Db 1644 TCGCCCAACGCGAGACCATTTTAAAGAAATAACATCTTTCAAGCCACGTAACCCCATGGA 1703  
Qy 1861 cctctcagaagacccctctgagcaactggaactatcttccaaagtccaaagatccaggtc 1920  
Db 1704 CCGCTCAGAGAGACCTCTGAGCAACTGGAATATCTTTCCAGAGTCCAGGATTTCCAGGTT 1763  
Qy 1921 gaatacaagaacttcccaaaaaacaaagaacgaatctgtatcctctatcaatctgctcc 1980  
Db 1764 GAATTAAGAACTTCCCAAAAAACAAAGAAAGAAATTTGTATCTATCAATTTGCTCC 1823

---

Qy 1981 tctcaagccacctctgaatcaagcgaatgtaatcgcgaagaatggtgaagtcctgcataatg 2040  
Db 1824 TCTCAGGACACCTCTGATCAGCCATGATGCGCAAGGATGTGGAGTGCCTCATGATATG 1883  
Qy 2041 gctgcgtgaacactctaaagtctgctgctggaatggaacaaagaatcagaatgca 2100  
Db 1884 GCTGCGGTGAACATCTTAAAGTGTCTGTGAGTTGGACCAACAAAGATACAGATGCCA 1943  
Qy 2101 agaacagaagaaggaacaaatgctgtgtgtggaagtgagcgaaccttcttgccatgaa 2160  
Db 1944 AGAACAGGAACGGACCAATGTCTGTGTGGAGTGTCTGAACCTTTTGTGGCATGAA 2003  
Qy 2161 ccaatlaaatacccaacataatactgaataaacttgaataacttgcattgaataatltggaat 2220  
Db 2004 CCAATTTAAATCCCAACATATATAGTAATATCTGAATATCTGAATCTGTTGAATTTGGAAT 2063  
Qy 2221 tcttgataacctcagtggtgcgaagaacagctgtggtlaaagatgtgtgagcagcaag 2280  
Db 2064 TTTCTGATACCTCCAGTGGCGCGAGAGACAGGTGGGTAAGATGTGGCAGCAGCGGG 2123  
Qy 2281 aagacaagaacaaacaaagaagcgtgtgtgcgcgtggaactgtgtggtgtgtgtgtgt 2340  
Db 2124 AAGACAACAGAAACACAAAGAGGCGCTGTGCGCGCTGAGACTGTGCTGGGTTGTGTGT 2183  
Qy 2341 gatgacacactcgttgaaactgtgctcctacgcaatagcagctgtgtgtggaagaag 2400  
Db 2184 GATGGCCACTCGGTGACCTGGCGGTCCCTACGCAATAGACGCTGTGGGAAAGAG 2243  
Qy 2401 gctgcgaagcagctgtgtctcccggaacacagcagaatccacacccctgggacactcgt 2460  
Db 2244 GCTGCCAGCAGCACTGTGTCTCCGCGACACAGCATCCACACCTTGCGGCACTCCGT 2303  
Qy 2461 gtttgtcttttttccctgtgtgaagaagaacagcagccctctcaagctgtg 2520  
Db 2304 GTTGTGCTTTTTCCTCTGTTGAAGAAAGAGGAGCAGACCCCTTCTAAGCTGG 2263  
Qy 2521 ctcaatcagaacatctggaacaaaccttggaacgaatgcaagaagaagcctctgacg 2580  
Db 2364 CTCACTCAGACATTTGGAGCAAAACCTTGACAGCCATGCAAGAGAGGCTTTGACCG 2423  
Qy 2581 gccccaagcctaaagaacacagaagaataaagtcttccactacagcgtgacccaact 2640  
Db 2424 GCCCGAGCTTAAGACAGCAGAGAAATCAATGCTTCTACTACAGCGGACCCAACTT 2483  
Qy 2641 tcttagtggtcgaagcgcacacacactcctgcagttacccaacacacacacactgtctc 2700  
Db 2484 TTTCTAGTGTCCACGGCCACACACCTCTGACGTACCCACACATCACACACTGCTTCT 2543  
Qy 2701 ctcccaagtgatctgtatctcttagttcatatcttcttgaatgtaataactat 2760  
Db 2544 CTTCCACACAGTACTGTATCTTATCTTGTTCATTAATTTCTTTGATTAATATACACTAT 2603  
Qy 2761 ataaaaatttcatcttgaataatctcgaatgtatctagttaaataagaacagtttgaac 2820  
Db 2604 ATAAATTTTCAATTTAGAAATTTCTCAATGTATCTAGTTAAATPAGACAGTTTGAAC 2663  
Qy 2821 tctgtgagagctgacttcatcaataatcaacgcgaagaagatcatatcatgtgtatgtg 2880  
Db 2664 TTTCTGTAGACTCTTATCAATTAATCAACGCAAAATATATCTCAATGTGTATGTG 2723  
Qy 2881 gttagaacatttcatcttcatctgaactaacccagaagagtttcagtgatgcaaatgtgtgc 2940  
Db 2724 GTTAGACATTTTATTTATTTCAATTTGACTTAACCCAGACAGTTTCACTGATGCAATTTGTGTC 2783  
Qy 2941 cctctgtctcagctgaacagctcctggaacttcaaaaaaccttgaataagttctcccaagt 3000  
Db 2784 CCGTGTGTTACGCTGAAGAGCTCGTGCAGCTTTCAAAAACCTGAATTAAGTCTCCACAGT 2843  
Qy 3001 tgtataaatgtgacaatttaagaatlttaaaacttgaatgataatctgtgttccatttca 3060  
Db 2844 TGTATTAATTTGAGCAATTTAAGAAATTTTAAACTTTAGATGATCATTTTGTTCATTTTAA 2903





QY	901	ctgcgcagagagcctcgaagtgaatbgyaagaaatccgaagaagaatctcaataatct	960
Db	901	ctgcgcagagagcctcgaagtgaatbgyaagaaatccgaagaagaatctcaataatct	960
QY	961	gaataaagtcgaagtgtttggagattgcactaaacggaacttcctgtgaatttcgaagt	1020
Db	961	gaataaagtcgaagtgtttggagattgcactaaacggaacttcctgtgaatttcgaagt	1020
QY	1021	gcccgaggagatbgycccccacacatgaagaactctgttaccaaagtctcgtttbgyag	1080
Db	1021	gcccgaggagatbgycccccacacacatgaagaactctgttaccaaagtctcgtttbgyag	1080
QY	1081	ctttgtggyggaagtgaaagyaagaatlttcaagaanaatbgyccgaatct	1140
Db	1081	ctttgtggyggaagtgaaagyaagaatlttcaagaanaatbgyccgaatct	1140
QY	1141	gtctctgagagcgtgaagaagtctacccctcgtccgtcgaattgaacggtaaagctbga	1200
Db	1141	gtctctgagagcgtgaagaagtctacccctcgtccgtcgaattgaacggtaaagctbga	1200
QY	1201	atcaaaagaagaaacaaacccatagltcaagccaacacaagcccgaaatatbgyccagag	1260
Db	1201	atcaaaagaagaaacaaacccatagltcaagccaacacaagcccgaaatatbgyccagag	1260
QY	1261	atcaatccgaattagcgaactbgycccgatltcagcagycaaaaaagagaaagagccagag	1320
Db	1261	atcaatccgaattagcgaactbgycccgatltcagcagycaaaaaagagaaagagccagag	1320
QY	1321	tacacgctcctcaaaagcggaagcctcccgcgcgcaaggaattgtgtatcgaagtgaag	1380
Db	1321	tacacgctcctcaaaagcggaagcctcccgcgcgcaaggaattgtgtatcgaagtgaag	1380
QY	1381	gttgaaacccaacatgtgcaagaagaagcgagccacaagaaggtbgyccagcgcaatga	1440
Db	1381	gttgaaacccaacatgtgcaagaagaagcgagccacaagaaggtbgyccagcgcaatga	1440
QY	1441	gccgagaaacatbgtcgtgaatccttgbtlttcaaaigtccgcagcgagccacccaaccc	1500
Db	1441	gccgagaaacatbgtcgtgaatccttgbtlttcaaaigtccgcagcgagccacccaaccc	1500
QY	1501	gcatcctaagtcagaaggaagaacacccataagaaccccgaggatbgyaagaagaataac	1560
Db	1501	gcatcctaagtcagaaggaagaacacccataagaaccccgaggatbgyaagaagaataac	1560
QY	1561	ctttctgaacctbgyctctbgyggaatgaagaatbgygacagtaataagaagataagatcag	1620
Db	1561	ctttctgaacctbgyctctbgyggaatgaagaatbgygacagtaataagaagataagatcag	1620
QY	1621	atgacctatctaagltcaacagcagctbgyctctbgtatcttcccaatbgttcccgagatc	1680
Db	1621	atgacctatctaagltcaacagcagctbgyctctbgtatcttcccaatbgttcccgagatc	1680
QY	1681	gcccaagctbtagagatltagtcaagaagatcacaccaagaatlttcaagagcgactcgcg	1740
Db	1681	gcccaagctbtagagatltagtcaagaagatcacaccaagaatlttcaagagcgactcgcg	1740
QY	1741	aatcctcgaagggcgaaggaactgcataagatagcccgagatbgttgtatbgygggaccc	1800
Db	1741	aatcctcgaagggcgaaggaactgcataagatagcccgagatbgttgtatbgygggaccc	1800
QY	1801	tcgcaccaagccgagacatlttcaagaataacatctltcagcgcaacgttacccaatga	1860
Db	1801	tcgcaccaagccgagacatlttcaagaataacatctltcagcgcaacgttacccaatga	1860
QY	1861	cctctcagagacccctctbgaacatbgyacatalttccaagaatccaggaatltcgaat	1920
Db	1861	cctctcagagacccctctbgaacatbgyacatalttccaagaatccaggaatltcgaat	1920
QY	1921	gaatacaagaacttccccaaaaacaagaagcaatltgtatctctatcaatgtctcc	1980
Db	1921	gaatacaagaacttccccaaaaacaagaagcaatltgtatctctatcaatgtctcc	1980

QY	1981	ttctaacgacctctgataccaacgtgtatctcggaagatgtggaatccctgcacatgata	tg	2040
Db	1981	ttctagaccacctctgataccaacgtgtatctcggaagatgtggaatccctgcacatgata	tg	2040
QY	2041	gctgcgcgcgaacatcttaagaattgctgtctgcgaattgcacaacaagatcacagaatgc	a	2100
Db	2041	gctgcgcgcgaacatcttaagaattgctgtctgcgaattgcacaacaagatcacagaatgc	a	2100
QY	2101	agaacaggaanaacggaaccaatgtctgtgtgtggaaggtgtctgaacctttctgtgcata	gaa	2160
Db	2101	agaacaggaanaacggaaccaatgtctgtgtgtggaaggtgtctgaacctttctgtgcata	gaa	2160
QY	2161	ccattataaatacccacaataatacttgaaataacgcgaacgccttttaaaatttgat	at	2220
Db	2161	ccattataaatacccacaataatacttgaaataacgcgaacgccttttaaaatttgat	at	2220
QY	2221	ttctgataacctcagatggccgcgaagacacagtgtagtaagaatgttggaacagcaag	g	2280
Db	2221	ttctgataacctcagatggccgcgaagacacagtgtagtaagaatgttggaacagcaag	g	2280
QY	2281	aagaacaagaanaacaacaaggagcgagctgtgtgcgagctgtgacgtgtgtgtgtgt	gt	2340
Db	2281	aagaacaagaanaacaacaaggagcgagctgtgtgcgagctgtgacgtgtgtgtgtgt	gt	2340
QY	2341	gatyggccaactcgtgtgacctgtgcgtgtccctacgcgaatagagctgtccgtgtgggaagaag	g	2400
Db	2341	gatyggccaactcgtgtgacctgtgcgtgtccctacgcgaatagagctgtccgtgtgggaagaag	g	2400
QY	2401	gctgcaccaagccaagctgtgtctccgcggagacacagagatctcacacctgtgggaacctcgt	g	2460
Db	2401	gctgcaccaagccaagctgtgtctccgcggagacacagagatctcacacctgtgggaacctcgt	g	2460
QY	2461	gtttgtgtcttttttttccctgtgtgtgaaagaanaacggaacgacacaccttctcaagctgtg	g	2520
Db	2461	gtttgtgtcttttttttccctgtgtgtgaaagaanaacggaacgacacaccttctcaagctgtg	g	2520
QY	2521	ctcaactcaagaacatctgtggagacaaacctgtgacagcatalgcagagaagagccctltgacg	g	2580
Db	2521	ctcaactcaagaacatctgtggagacaaacctgtgacagcatalgcagagaagagccctltgacg	g	2580
QY	2581	gcccagagctaaagaacacagagaaatcaaatgtcttctactaagcgtgaaccaact	tt	2640
Db	2581	gcccagagctaaagaacacagagaaatcaaatgtcttctactaagcgtgaaccaact	tt	2640
QY	2641	ttctagtgatgccaagcgcccaacacacctctgtcagtaaccacaaccaatcaacatgcttct	tt	2700
Db	2641	ttctagtgatgccaagcgcccaacacacctctgtcagtaaccacaaccaatcaacatgcttct	tt	2700
QY	2701	cttccaacatgatctgtatcttctagtctcaatattcttctttgatgatatagacata	t	2760
Db	2701	cttccaacatgatctgtatcttctagtctcaatattcttctttgatgatatagacata	t	2760
QY	2761	ataaatttcaatttggaaatttctcaattgtatctcaagttaaataagacaadtgtggaaac	g	2820
Db	2761	ataaatttcaatttggaaatttctcaattgtatctcaagttaaataagacaadtgtggaaac	g	2820
QY	2821	ttgtcgtagacgtacgtttatcatatattctaaacgcgaamaagatcatatcattgtatgtg	g	2880
Db	2821	ttgtcgtagacgtacgtttatcatatattctaaacgcgaamaagatcatatcattgtatgtg	g	2880
QY	2881	gttagacaatttatttatttatttgaactaaaccogaagctttcagatgtgcaaatgtgtgc	g	2940
Db	2881	gttagacaatttatttatttatttgaactaaaccogaagctttcagatgtgcaaatgtgtgc	g	2940
QY	2941	ccctctggttcaagctgaaacgctccctgtgactttcaaaaccttgaaataagttcccaacgt	g	3000
Db	2941	ccctctggttcaagctgaaacgctccctgtgactttcaaaaccttgaaataagttcccaacgt	g	3000
QY	3001	tgataaattggacaatttggagaatttgaacttgaatcatctgtgtgttcaattt	t	3060
Db	3001	tgataaattggacaatttggagaatttgaacttgaatcatctgtgtgttcaattt	t	3060
QY	3061	tttaatttatttatttlttatttgaacaaacggaactttaaagaaacttgaacttgcctctgtttaa	a	3120

22	150.8	4.3	160979	52	AC055879	Homo sapi	
23	149.6	4.3	3173	34	AK001576	Homo sapi	
C	24	149	4.2	332811	72	AL1390158	Homo sapi
25	123.4	3.8	194407	54	AC068845	Homo sapi	
26	125.2	3.6	2008	34	AC023314	Homo sapi	
27	119	3.4	2847	69	HSRNTAFPI	X83973 H. sapiens m	
28	110.2	3.1	5237	52	DROSTAFEN	M69111 D. melanogus	
29	110.2	3.1	61267	28	AC004336	AC004336 Drosophil	
C	30	110.2	3.1	120625	38	AC017563	Drosophil
C	31	110.2	3.1	267271	28	AE003800	AE003800 Drosophil
C	32	105.8	3.0	8409	29	AF225924	AF225924 Drosophil
33	103	2.9	8002	29	AF225925	AF225925 Musc dom	
C	34	85.4	2.4	171591	50	AC027018	AC027018 Homo sapi
C	35	74	2.1	7218	81	166494	166494 Sequence 14
C	36	68	1.9	7218	81	166494	166494 Sequence 14
37	65.4	1.9	60009	48	AC023436	AC023436 Homo sapi	
38	63	1.8	180968	62	AC078996	AC078996 Mus muscu	
C	39	63	1.8	258550	35	AC010758	AC010758 Homo sapi
C	40	62.6	1.8	42059	29	CELF5544	CELF5544 Caenorhabd
C	41	62.4	1.8	141892	47	AC023197	AC023197 Mus muscu
42	62.2	1.8	89665	69	HSJ340H1	AL109755 Homo sapi	
43	61.4	1.8	77127	36	AC011801	AC011801 Homo sapi	
44	61.2	1.7	92709	37	AC015716	AC015716 Homo sapi	
C	45	61	1.7	80725	34	AL137078	AL137078 Homo DNA

## ALIGNMENTS

RESULT	1
LOCUS	AF061939
DEFINITION	AF061939 3506 bp mRNA PRI 04-MAR-1999
ACCESSION	Homo sapiens staufen protein (STAU) mRNA, alternatively spliced, complete cds.
VERSION	AF061939
KEYWORDS	AF061939.1 GI:4335948
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Matayola, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	1 (bases 1 to 3506)
TITLE	Wickham, L., Duchateau, T., Luo, M., Nabi, I. R. and Desgroselliers, L. Mammalian staufen is a double-stranded RNA and tubulin binding protein which localizes to the rough endoplasmic reticulum
JOURNAL	Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
MEDLINE	99147057
REFERENCE	2 (bases 1 to 3506)
AUTHORS	Desgroselliers, L., Wickham, L. and Luo, M.
TITLE	Direct Submission
JOURNAL	Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900
FEATURES	Edouard Montperril, Montreal, Quebec H3T 1J4, Canada
source	Location/Qualifiers
	1..3506
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="20"
	/map="20q13.1"
	1..3506
	/gene="STAU"
	409..2142
	/gene="STAU"
	/function="RNA binding protein"
	/note="alternatively spliced transcript"
	/codon_start=1
	/product="staufen protein"
	/protein_id="AA017531.1"
	/db_xref="GI:4335949"
	/translation="MSOVQVQNSPALGSGQLLNKNSLIQPIWISPTSSIP
	ENMRPIONSLPSSATSTSAAEISTTPYELNALCKLKKPKATKVDYTRKOST
	YINMRGATPRITFPYVPLLYVELSGVGOOFNGKTRQAKHDAKARITII
	ONEPLPERLEWSESEENLNKSEISQVEFIALKRLNPVNEEVARSEGPHMNEV
	KVSGVEGVDEGEGSKRIKKNAAIVALEELKLPPLPAVRVPRIRIKKPTVPC

[illegible]

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2001, 08:45:50 ; Search time 24649.1 Seconds  
(without alignments)  
727.932 Million cell updates/sec

Title: US-09-316-048-5  
Perfect score: 3506  
Sequence: 1 actctctgcggctgcggg.....taatactaaaaaaaaa 3506

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_cm:\*  
4: gb\_ov:\*  
5: gb\_ph:\*  
6: gb\_pi1:\*  
7: gb\_pi2:\*  
8: gb\_pi1:\*  
9: gb\_pi2:\*  
10: gb\_pi3:\*  
11: em\_fun:\*  
12: em\_hum1:\*  
13: em\_hum2:\*  
14: em\_in:\*  
15: em\_om:\*  
16: em\_or:\*  
17: em\_ov:\*  
18: em\_pat:\*  
19: em\_ph:\*  
20: em\_pi:\*  
21: em\_ro:\*  
22: em\_sts:\*  
23: em\_sy:\*  
24: em\_un:\*  
25: em\_vl:\*  
26: gb\_hcg1:\*  
27: gb\_hcg2:\*  
28: gb\_hcg3:\*  
29: gb\_hcg4:\*  
30: em\_ba1:\*  
31: em\_ba2:\*  
32: em\_ba3:\*  
33: em\_hum3:\*  
34: em\_hum4:\*  
35: gb\_pi4:\*  
36: gb\_hcg3:\*  
37: gb\_hcg4:\*  
38: gb\_hcg5:\*  
39: gb\_hcg6:\*  
40: em\_hcg1:\*  
41: em\_hcg2:\*  
42: em\_hcg3:\*  
43: em\_hum5:\*

44: gb\_pi3:\*  
45: gb\_pi5:\*  
46: gb\_hcg8:\*  
47: gb\_hcg9:\*  
48: gb\_hcg10:\*  
49: gb\_hcg11:\*  
50: gb\_hcg12:\*  
51: gb\_hcg13:\*  
52: gb\_hcg14:\*  
53: gb\_pi3:\*  
54: gb\_hcg15:\*  
55: gb\_hcg16:\*  
56: gb\_hcg17:\*  
57: em\_hcg4:\*  
58: em\_hcg5:\*  
59: em\_hcg6:\*  
60: em\_hcg7:\*  
61: em\_hcg8:\*  
62: gb\_hcg18:\*  
63: gb\_hcg19:\*  
64: gb\_ba3:\*  
65: em\_hcg8:\*  
66: em\_hcg9:\*  
67: em\_hcg10:\*  
68: gb\_pi6:\*  
69: gb\_pi7:\*  
70: gb\_hcg20:\*  
71: gb\_hcg21:\*  
72: gb\_hcg22:\*  
73: gb\_hcg23:\*  
74: gb\_ro:\*  
75: gb\_sts1:\*  
76: gb\_sts2:\*  
77: gb\_sy:\*  
78: gb\_un:\*  
79: gb\_vl1:\*  
80: gb\_vl2:\*  
81: gb\_pat1:\*  
82: gb\_pat2:\*  
83: em\_hcg0:\*  
84: gb\_hcg24:\*  
85: gb\_pi8:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3502.8	99.9	3506	10 AF061939	AF061939 Homo sapi
2	2958.8	84.4	3349	10 AF061941	AF061941 Homo sapi
3	2914.8	83.1	3217	10 AF061938	AF061938 Homo sapi
4	2896	82.6	3142	10 AF061940	AF061940 Homo sapi
5	2848.8	81.3	3194	34 AK025519	AK025519 Homo sapi
6	2799.4	79.8	3066	68 HSA132258	AJ132258 Homo sapi
7	1357.4	38.7	113836	34 AL133174	AL133174 Human DNA
8	1282.8	36.6	3041	74 AF290989	AF290989 Rattus no
9	1256	35.8	2860	74 AF061942	AF061942 Mus muscu
10	1254.8	35.8	3059	74 AF227200	AF227200 Rattus no
11	705	20.1	157803	35 AC011492	AC011492 Homo sapi
12	553.2	15.8	194407	54 AC068845	AC068845 Homo sapi
13	383.2	10.9	192597	70 AL158850	AL158850 Homo sapi
14	380.2	10.8	385	75 G30939	G30939 DRESSO Huma
15	291	8.3	160979	52 AC055879	AC055879 Homo sapi
16	287	8.2	360	74 RMO010200	AJ010200 Rattus no
17	283.6	8.1	225016	52 AC063967	AC063967 Mus muscu
18	276.8	7.9	2968	34 AK002152	AK002152 Homo sapi
19	276.8	7.9	4058	68 HSA9062	Y19062 Homo sapien
20	220.2	6.3	263	75 G22107	G22107 human STR W
21	193.6	5.5	163253	68 HSB140E4	AL121877 Human DNA



Db 145 TTGCTTTGATGTTGTCTTACTGTACACAGAGTGTCGACACTGAGCCTCTGCGTGTGCT 86  
Qy 3080 ccgatatggaacacctggttagccctgcgagttcaagtactgcttccatcatctgtttacgct 3139  
|||||  
Db 85 CCGATGGAAGAGCTGTGTAGCCCTGCAGATTAAAGTACTGCTTCATTCATTTGTTAGGCT 26  
Qy 3140 ggaatttttcccatcggaatgta 3164  
|||||  
Db 25 GGAATTTTCTCCCAAGATGTA 1

## RESULT 15

RNO010200 360 bp mRNA ROD 22-JAN-1999  
LOCUS Rattus norvegicus mRNA for Staufen protein, partial.  
DEFINITION AJ010200  
ACCESSION AJ010200.1 GI:4138427  
VERSION staufen gene.  
KEYWORDS Norway rat.  
SOURCE Rattus norvegicus  
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 360)  
AUTHORS Kiebler M.A.  
TITLE Direct Submission  
JOURNAL Submitted (16-AUG-1998) Kiebler M.A., Cell Biology, EMBL,  
Meyerhofstr. 1, D-69117 Heidelberg, GERMANY  
REFERENCE 2 (bases 1 to 360)  
AUTHORS Kiebler M.A., Hemraj, I., Verkade, P., Kohrman, M., Fortes, P.,  
Marion, R.M., Orth, J., and Dotti, C.G.  
TITLE The mammalian staufen protein localizes to the somatodendritic  
domain of cultured hippocampal neurons: Implications for its  
involvement in mRNA transport  
J. Neurosci. 19 (1), 288-297 (1999)

JOURNAL MEDLINE 99088098  
FEATURES  
source Location/Qualifiers  
1..360  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone\_lib="Rat brain"  
/dev\_stage="embryonic"  
1..119  
/gene="staufen"  
<1..>119  
/gene="staufen"  
/codon\_start=2  
/evidence=experimental  
/product="Staufen protein"  
/protein\_id="CAA09037.1"  
/db\_xref="GI:4138428"  
/translation="YFYPPVPVPLLYOVELSVGGQGFNGKGMKRPVKHDATA"

BASE COUNT 109 a 70 c 105 g 76 t  
ORIGIN

Query Match 8.9%; Score 287; DB 74; Length 360;

Best Local Similarity 87.5%; Pred. No. 2.4e-56;

Matches 314; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 465 taacttaccattccagttccagttcaacttactatataagtggaacttctgttgagaga 524  
|||||  
Db 2 TACTTTTACCAATTCAGTTCACACCGTCTACTCTACCAAGTTGAGCTCTCCGTGAGAGA 61  
Qy 525 caagcaatttaatggaagaagaagacaagcgtctggaacacgagtgctgtgcca 584  
|||||  
Db 62 CACGAGTTTATGAGGAGGAGATGAGACCACTGTGAACATGATGCCACTGCCCGA 121  
Qy 585 gcgttgagagatccctgcagaaatgagccctgcagagagagctggaagtgaatggaagaa 644  
|||||  
Db 122 GCCCTGAGGACTCTGCAGAGTGAAGCCCTTACACAGAGAGGCTAGAGTAATGGAAGAA 181  
Qy 645 tccgaagaagaataatcctcaataatcgaataagtaagtgattgagattgcaactaa 704  
|||||

Db 182 TCAGACAGAGAAAACCTTAAATCAAGAAATAAGCCAAGTGTTCAGATTGCCCTGAG 241  
Qy 705 cgaacttgctgtgaaatttcgaggtgcccggagagtgccccacccacatgaagaac 764  
|||||  
Db 242 CGGAATTTGGCTGTGAATTTTGAAGGTGCGCGGAGAGAGTGCGCGCCACACATGAAGAAC 301  
Qy 765 ttgtgaccaaaggttctggttgaggagttgtgtggggaaggtgaagggaagaagaa 823  
|||||  
Db 302 TTGTGACACAGAGGTTTCAGTTGGGAGAGTTGTAGGGAGAGAGAAAGCAAGAA 360

Search completed: April 5, 2001, 08:45:50  
Job time: 78395 sec

```

Db 60919 GATATTTTATTTGGCTGTGAAAAAGAAATGGAAGACAGAAAGATTAACCTGCTCAG 60860
QY 2237 tcagacacattgaggacaacccctgagcagcattgcagagagagcccttgacggcccc 2296
Db 60859 ACTGGCTGACTCAAC -ACCTTGTCTTAACATGCTGGAGAAAGCCCTTAGCTGCTC 60801
QY 2297 agagctaaagacagcagagaaataagcttcctcactcagcgtgagcccaatttcta 2356
Db 60800 AAGCTGAAACACCGGGGAAATTAACGTCTTCTACCTGTTACAAATTAACCTTCTTA 60741
QY 2357 gctgcaagagcccccacccctcctgcaqtacacacacacacacacacacacacacac 2416
Db 60740 GCATGCTAGG -CCAGTCACTTACTGCACTTACCCACACATTAACACCTGCTTCTTCT 60682
QY 2417 aacagtgatcgtatctctagtttcaatttcttcttgatgatgatagaactataaa 2476
Db 60681 AATAGTGAATATATTTTATTTATTTA -TTTCTTTGATTGAATGATGCATATGAA 60623
QY 2477 atttcatctggaatttctcaattgatactagtttaataagacagtttgaaactgtc 2536
Db 60622 ATTTCATTTGAAAGTTTCTCAGTTGTATCTAGTTACATGATACAGTTCAGAAACCTGTC 60563
QY 2537 tgagactgacttatcaataatcaatcaacacacacacacacacacacacacacacac 2596
Db 60562 CCAGACTGACTTTATCAGAAATCTAATGCAATGATCATATCCATGATAGCTGTTGT 60503
QY 2597 acatttcatctcaattgataacacagacagttcagtgatgataaattggtgccccct 2656
Db 60502 ACATTTCTTTCTACTGATGATACCAAGAACCATTTTCAGTATGCAAAATCATGACCTCT 60443
QY 2657 gtttgcagctgaagacagctcctgactt-tcaaaaccttgaataagctcccaagtgt 2715
Db 60442 GGTTTAGCTGAAACCCCTGGACTTCTCAGAAACCTTATCAATTTTCACATTTGTA 60383
QY 2716 taatgtgacaattgagaatttgaatttgaattgagatgattggttcaatttcaatt 2775
Db 60382 CAAATTTGGCATTTTATGGAATTTCAAACTTCA ---ATCATTTGGTCTCTTTATATTTT 60327
QY 2776 atttcatcttctgtaataatca--aacaggaactaaatgaacttgcctcgttttaag 2833
Db 60326 ATTTTATTTTGGTGTCTGTAATGGGACCTTAACCAATTAACCTTGTATCTTTTAAAG 60267
QY 2834 attatataaaacattgtaataatcaataatgataatgagctcctgagactgaacttca 2893
Db 60266 ATTATTAATAAATCTTACGTGATATGATACATATAGCTTTCAAGACCCACAGCTTCACTATG 60207
QY 2894 ctacagatatgatactcatalgtatgatacatalaataaacctgagagtgatttcc-agagtg 2952
Db 60206 CTGTAGATATTTGTCTCTGTAGTAGTACATATGACATTTGCAAGTATTTTCCCTTGAAGTG 60147
QY 2953 tcgatactgtaataatcaatc 2972
Db 60146 TTTATACCTTATTTTCTC 60127

```

## RESULT 14

```

LOCUS 385 bp DNA STS 29-SEP-1998
DEFINITION DRES50 Human DRES Homo sapiens STS genomic clone 22368 3', sequence
tagged site.
ACCESSION G30939
VERSION G30939
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 385)
AUTHORS Banfi,S., Borsani,G., Rossi,E., Bernard,L., Guffanti,A.,
Rubio,L.F., Marchitelli,A., Giglio,S., Coluccia,E., Zollo,M.,
Zuffardi,O. and Ballabio,A.
TITLE Identification and mapping human cDNAs homologous to Drosophila

```

JOURNAL  
MEDLINE  
COMMENT  
mutant genes through EST database searching  
Nat. Genet. 13 (2), 167-174 (1996)  
96225443

Contact: Ballabio Andrea  
Telethon Institute of Genetics and Medicine, TIGEM  
Telethon Institute of Genetics and Medicine, TIGEM  
via Olgettina 58, 20132 Milano, Italy  
Tel: +39-2-215601  
Fax: +39-2-21560220  
Email: dres@tigem.it  
Primer A: GTAAACAATGAATGAGACAG  
Primer B: GGCGTGAAGAATGACCTAC  
STS size: 157  
PCR profile:

Presoak: 94 degrees C for 2 minutes  
Denaturation: 94 degrees C for 45 seconds  
Annealing: 57 degrees C for 45 seconds  
Polymerization: 72 degrees C for 45 seconds  
PCR Cycles: 35  
Thermal Cycler: MJR PTC-200

Protocol:  
Template: 25 ng  
Primer: each 1 uM  
dNTPs: each 220 uM  
Tag Polymerase: 0.025 units/ul  
Total Vol: 15 ul

Buffer:  
MgCl2: 1.5 mM  
KCl: 50 mM  
Tris-HCl: 10 mM  
pH: 8.3

FEATURES  
source  
The 5' end of the cDNA clone 22368 is homologous to the staufer  
Drosophila gene product. EST from which STS is derived: T89105.  
Forward primer is T0-1233; Reverse primer is T0-1234.  
location/Qualifiers

1..385  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="20, 12.10 cR from GCT10F11 (10d >3.0)"  
/clone="22368"  
/clone\_11b="Human DRES"  
STS  
primer\_bind  
29..185  
primer\_bind  
29..49  
complement(165..185)  
BASE COUNT 135 a 79 c 71 g 100 t  
ORIGIN

Query Match 11.8%; Score 380.2; DB 75; Length 385;  
Best Local Similarity 99.2%; Pred. No. 4.4e-78;  
Matches 382; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 2780 ttttttttgaagaaacagacttaataatgaacttgaactcgtttaaagattatt 2839
Db 385 TTTATTTGTTATGCAACAGGACTTAATGACATTTGATCTGTTTAAAGATTATT 326
QY 2840 aaaaacattgttactatatacatalgtctctgagagacttgaacttcaactacacag 2899
Db 325 AAAAAACATTGTATCTATATACATATGAGCTTTGAGAGACTTACGTTTCACTACACAG 266
QY 2900 gatagatctccatgtagtcacataaaccctgagagtgatttccagagtgctgatac 2959
Db 265 GATATGATCTCCATGTAGTCACATATTAACCTGCAGAGTGATTTTCCAGAGTGTGATAC 206
QY 2960 tgttaattacatctcattagggctgaaagaaatgacacgttctgtatcacagctgtg 3019
Db 205 TGTATATACATCTCATTTAGGCTGAAAAGATGACCTACAGTTTCTGTATACAGCTG 146
QY 3020 ttgctttgatgtgtgttactgtacacagaagtgtgcaactgagagctcgtgtgtgt 3079

```

QY 2111 ggtctccagccagctggtctccgggagaccagcagatccacacctgggcaacctcg 2170  
 |||||||  
 Db 2946 GAGTGCACGACGACGCTGCTCTGGGGGACACAGATCCACGCCCTGGGACACTTCC 2887  
 |||||||  
 QY 2171 tgttggctcttttttccctctg--tgaaagaagaagcgacgaccccttcaagc 2228  
 |||||||  
 Db 2886 TGTGTGGTCTTTTTCCTCCGTGAAGAAAGTGAAGATGACCCCTTCTCAAGC 2827  
 |||||||  
 QY 2229 tggctcactcagacacattggagacaacctggagacgacatgacagagagccttga 2288  
 |||||||  
 Db 2826 TGGCTTA-----GTGGGACAAACCCCGACACACACAGAGAGAGGCTTCAA 2776  
 |||||||  
 QY 2289 ccggcccaagagctaaagaccagagaaaataatgcttctactcaagctgacccaa 2348  
 |||||||  
 Db 2775 CCAGCCCGACGAGCTAAAGACACAGAAAGTCAATGCTTCTATTACAGATGACCTAA 2716  
 |||||||  
 QY 2349 ctcttcagtggtgcagggcccaacccctctctgcaatccacacataccacgtctt 2408  
 |||||||  
 Db 2715 C--TTCTAGTGTGCGACGGCCCGCCACTTCTGCTGACACACATCACACTGCTT 2658  
 |||||||  
 QY 2409 tctctcacaagtgatctatcttctgattcatatttcttcttgatgatagcac 2468  
 |||||||  
 Db 2657 TCTCTTCCACAGATCTGATTCTTAGTTTCAATTATTCCTTTTATTGCTATGACAC 2598  
 |||||||  
 QY 2469 tatataaatlctcat 2484  
 |||||||  
 Db 2597 TATATAAAATTTCTCT 2582  
 |||||||

RESULT 13  
 AL158850/c  
 LOCUS  
 DEFINITION  
 AL158850  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

AL158850 192597 bp DNA HTG 06-SEP-2000  
 Homo sapiens chromosome 6 clone RP3-522B19, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*, 11 unordered pieces.  
 AL158850  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
 human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 192597)  
 Direct Submission  
 Submitted (05-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone  
 requests: clonerequest@sanger.ac.uk  
 On Aug 13, 2000 this sequence version replaced g1:9212635.  
 ----- Genome Center  
 Center: Sanger Centre  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquerry@sanger.ac.uk  
 ----- Project Information  
 Center project name: dj522B19  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Sequencing vector: M13; M7815; 12% of reads  
 Sequencing vector: plasmid; L08752; 87% of reads  
 Chemistry: Dye-terminator Big Dye; 77% of reads  
 Chemistry: Dye-terminator-amersham; 22% of reads  
 Chemistry: Dye-terminator Big Dye; 0% of reads  
 Consensus quality: 185625 bases at least Q40  
 Consensus quality: 188533 bases at least Q30  
 Consensus quality: 190005 bases at least Q20  
 Insert size: 191597; sum-of-contigs  
 Quality coverage: 5.71x in Q20 bases; sum-of-contigs Quality  
 coverage: 8.21x in Q20 bases; agarose-efp  
 -----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 11 contigs. The true order of the pieces

\* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 13837: contig of 13837 bp in length  
 \* 13838 13937: gap of 100 bp  
 \* 13938 17476: contig of 3539 bp in length  
 \* 17477 17576: gap of 100 bp  
 \* 17577 20005: contig of 2429 bp in length  
 \* 20006 20105: gap of 100 bp  
 \* 20106 32070: contig of 11965 bp in length  
 \* 32071 32170: gap of 100 bp  
 \* 32171 39772: contig of 7602 bp in length  
 \* 39773 39872: gap of 100 bp  
 \* 39873 43471: contig of 3599 bp in length  
 \* 43472 43571: gap of 100 bp  
 \* 43572 51795: contig of 8224 bp in length  
 \* 51796 51895: gap of 100 bp  
 \* 51896 178838: contig of 126943 bp in length  
 \* 178839 178938: gap of 100 bp  
 \* 178939 181919: contig of 2981 bp in length  
 \* 181920 182019: gap of 100 bp  
 \* 182020 184231: contig of 2212 bp in length  
 \* 184232 184331: gap of 100 bp  
 \* 184332 192597: contig of 8266 bp in length.  
 \* location/Qualifiers  
 \* 1. 192597  
 \* /organism="Homo sapiens"  
 \* /db\_xref="taxon:9606"  
 \* /chromosome="6"  
 \* /clone="RP3-522B19"  
 \* /clone\_1lb="RPC1-3"  
 \* 1. 13837  
 \* /note="assembly\_fragment:02383"  
 \* fragment\_chain:1"  
 \* 13938. 17476  
 \* /note="assembly\_fragment:01080"  
 \* fragment\_chain:1"  
 \* 17577. 20005  
 \* /note="assembly\_fragment:02434"  
 \* fragment\_chain:1"  
 \* 20106. 32070  
 \* /note="assembly\_fragment:03060"  
 \* fragment\_chain:2"  
 \* 32171. 39772  
 \* /note="assembly\_fragment:00945"  
 \* fragment\_chain:2"  
 \* 39873. 43471  
 \* /note="assembly\_fragment:00013"  
 \* fragment\_chain:2"  
 \* 43572. 51795  
 \* /note="assembly\_fragment:02225"  
 \* fragment\_chain:3"  
 \* 51896. 178838  
 \* /note="assembly\_fragment:01767"  
 \* fragment\_chain:3"  
 \* 178939. 181919  
 \* /note="assembly\_fragment:01637"  
 \* 182020. 184231  
 \* /note="assembly\_fragment:02408"  
 \* 184332. 192597  
 \* /note="assembly\_fragment:02960"  
 \* BASE COUNT 58324 a 39515 c 38313 g 53431 t 1014 others  
 ORIGIN

QY 2177 gttcttttccctctgtaagaagaagcagcacccttccaagctggtccac 2236

Query Match 11.9% Score 383.2; DB 70; Length 192597;  
 Best Local Similarity 73.2%; Pred. No. 1.7e-78;  
 Matches 586; Conservative 0; Mismatches 203; Indels 11; Gaps 7;



*	1856	3429:	contig of 1574 bp	in length
*	3430	3529:	gap of 100 bp	
*	3530	5346:	contig of 1817 bp	in length
*	5347	5446:	gap of 100 bp	
*	5447	7338:	contig of 1892 bp	in length
*	7339	7438:	gap of 100 bp	
*	7439	10297:	contig of 2859 bp	in length
*	10298	10397:	gap of 100 bp	
*	10398	13590:	contig of 3193 bp	in length
*	13591	13690:	gap of 100 bp	
*	13691	16293:	contig of 2603 bp	in length
*	16294	16393:	gap of 100 bp	
*	16394	19065:	contig of 2672 bp	in length
*	19066	19165:	gap of 100 bp	
*	19166	22713:	contig of 3548 bp	in length
*	22714	22813:	gap of 100 bp	
*	22814	27165:	contig of 4352 bp	in length
*	27166	27265:	gap of 100 bp	
*	27266	32377:	contig of 5112 bp	in length
*	32378	32477:	gap of 100 bp	
*	32478	36572:	contig of 4095 bp	in length
*	36573	36672:	gap of 100 bp	
*	36673	41657:	contig of 4985 bp	in length
*	41658	41757:	gap of 100 bp	
*	41758	47024:	contig of 5267 bp	in length
*	47025	47124:	gap of 100 bp	
*	47125	52938:	contig of 5814 bp	in length
*	52939	53038:	gap of 100 bp	
*	53039	59129:	contig of 6091 bp	in length
*	59130	59229:	gap of 100 bp	
*	59230	63896:	contig of 6667 bp	in length
*	63897	65996:	gap of 100 bp	
*	65997	89427:	contig of 2331 bp	in length
*	89428	89527:	gap of 100 bp	
*	89528	95785:	contig of 6258 bp	in length
*	95786	95885:	gap of 100 bp	
*	95886	102762:	contig of 6877 bp	in length
*	102763	102862:	gap of 100 bp	
*	102863	110899:	contig of 8037 bp	in length
*	110900	110999:	gap of 100 bp	
*	110901	121002:	contig of 10003 bp	in length
*	121003	121102:	gap of 100 bp	
*	121103	133048:	contig of 11946 bp	in length
*	133049	133148:	gap of 100 bp	
*	133149	145755:	contig of 12607 bp	in length
*	145756	145855:	gap of 100 bp	
*	145856	165006:	contig of 19151 bp	in length
*	165007	165106:	gap of 100 bp	
*	165107	188243:	contig of 21337 bp	in length
*	188244	188343:	gap of 100 bp	
*	188344	194407:	contig of 6064 bp	in length.
FEATURES				
SOURCE				
Location/Qualifiers				
1..194407				
"organism="Homo sapiens"				
"db_xref="taxon:9606"				
"chromosome="19"				
"map="19"				
"clone="RP11-73711"				
"clone_id="RPCT-11 Human Male BAC"				
1..1755				
"note="assembly_fragment"				
misc_feature				
1856..3429				
"note="assembly_fragment"				
misc_feature				
3530..5346				
"note="assembly_fragment"				
misc_feature				
5447..7338				
"note="assembly_fragment"				
misc_feature				
7439..10297				
"note="assembly_fragment"				
misc_feature				
10398..13590				
"note="assembly_fragment"				
misc_feature				
13691..16293				

	misc_feature	16394..19065	/note="assembly_fragment"
	misc_feature	19166..22713	/note="assembly_fragment"
	misc_feature	22814..27165	/note="assembly_fragment"
	misc_feature	27266..32377	/note="assembly_fragment"
	misc_feature	32478..36572	/note="assembly_fragment"
	misc_feature	36673..41657	/note="assembly_fragment"
	misc_feature	41758..47024	/note="assembly_fragment"
	misc_feature	47125..52938	/note="assembly_fragment"
	misc_feature	53039..59129	/note="assembly_fragment"
	misc_feature	59230..65896	/note="assembly_fragment"
	misc_feature	65997..89427	/note="assembly_fragment"
	misc_feature	89528..95785	/note="assembly_fragment"
	misc_feature	95886..102762	/note="assembly_fragment"
	misc_feature	102863..110899	/note="assembly_fragment"
	misc_feature	111000..121002	/note="assembly_fragment"
	misc_feature	121103..133048	/note="assembly_fragment"
	misc_feature	133149..145755	/note="assembly_fragment"
	misc_feature	145856..165006	/note="assembly_fragment"
	misc_feature	165107..188243	/note="assembly_fragment"
	misc_feature	188344..194407	/note="assembly_fragment"
	misc_feature	clone_end:r7	
	vector_side:right		
BASE COUNT	50993 a	43965 c	46383 g 50458 t 2608 others
ORIGIN			
Query Match	17.2%	Score 553.2:	DB 54; Length 194407;
Best Local Similarity	90.4%	Pred. No.3.9e-118;	
Matches 665;	Conservative	0; Mismatches 48;	Indels 23; Gaps 6.
Oy	1752	gctgcgtgaacatcttaagttgcgtcgttccttgaccacaagaatacagatatcca	1811
Dd	3297	gctgcacatgacaccttaaaagtctgctgttgaattggacaccaaagttcacagattgccca	3238
Oy	1812	agaacagaagaacgcgaaccaatgctcgtgctgtgtagaggtgctgaaccttttcggccatgaa	1871
Dd	3237	AGAACAAGGAAATGATTCACATGTCTGTGTGGTGGAGAGTGCTGAACCTTTTCTGCCATGAA	3178
Oy	1872	ccattatataaatcccacaotatatatctgaaaactcgaacagcgttttgaaaaatttggnaat	1931
Dd	3177	CCATTATATAAATCCCAACATTAATACCTGAAAATATTAAAACTGCTTTGAAAAATTGGAAAT	3118
Oy	1932	tctctataacctcgaatggtgcgaagagacacggttggttaaaggatgtggtgcacgacgaag	1991
Dd	3117	TTCCTGATACCTTACTAGTGGCCGACAGG--ACGGTGGGTAAGAAGATGTGGCACACAGAGG	3060
Oy	1992	aagacaacagaanaacaaagaggcggtctgtgac--ggtcgaacgtgtcctgggttttgtt	2050
Dd	3059	AACACAACAG-----GGGAGGTGGCTGTGGCCGGCGCTGGACGTGTGCTGGGGTTGTGG	3007
Oy	2051	tgatgtccactgtgtgacctgtgcgttccctlaagcaatgacacatcctgcgttggggaagaag	2110
Dd	3006	TTAGGCACCTCCTGGTAGCCTGTGCTGTCCCTACGCAATACCACTGCTCTTGGGGAAGAAG	2947

```

Db 150307 C--TTTAGTGTGGCCAGGCCCCACACACTTCCTGACAGTGCCACACACACACTGCTT 150250
QY 2409 tctcttccacagtgatctgattcttaagtttaattcttcttgatgat----- 2463
Db 150249 TCTCTTCCAAACAGTATCTGATTTCTTAGTTTCAATTATTTCTTTTATTTGATGACAC 150190
QY 2464 ----- 2463
Db 150189 TATATATAATTTTCCTCTTCCTGCTTCCTTCCTTCCTTTCTTTCTTTCTTTCTTCT 150130
QY 2464 ----- 2463
Db 150129 TTCTTTCTTTCAAGACAAAGTCTCACTCTGTACCCAGGCTGAGTCAACGGCACAGTC 150070
QY 2464 ----- 2463
Db 150069 TTGGCTCCCTGCACACTTCTGCTCCAGGTTCAGCAATTTGTCTGCTCAGCTCCGAG 150010
QY 2464 ----- 2463
Db 150009 TAGCTGGGCCACACACCCAGTAGTTTTTTGTACTTTTAATAGAGCGGGTTTCACCAT 149950
QY 2464 ----- 2463
Db 149949 CTGGCCAGGCTGGTCTTGAATCTGTGACCTGTGATCCACCCGCTCGGCTCCCAAG 149890
QY 2464 ----- 2463
Db 149889 TGCTGGGATTACTGGCATYGGCCACCGCTGCCCGACACACTATATAAATTTTCATTGGA 149830
QY 2489 gaatttctcaatigtatctagtttaaatagacagtttggaactgtctgagactgactt 2548
Db 149829 GAGTTTCTCAAGTATCTAATTAATATGCGACAGTTTGAAGCTTTCTGAGACTGACTT 149770
QY 2549 tatcaataactcaacgcagaagaatcatccatgctgtagtctgagactgacttattt 2608
Db 149769 TATCAATAAATCTAAACCAAGATCATATTCACGCTGTATGTAGTATTTTATTTTATTT 149710
QY 2609 cattgactaacccagcagcttcaagtatgacgtaaatgtgtgaccttggttcagctgaa 2668
Db 149709 CATTGACTAACCCAGCACCACTTTCAGTATGCAAAATTACGTGCTTATGTTGCTGACTT 149650
QY 2669 acagtcctgactttcaaaaacttgaataagtcctccacagttgtataaatggaact 2728
Db 149649 ACAGTCTCGACTTTCAAAAACCTTGAAATGAAGTCGCCACAGTTGTAATAATTGACAAAT 149590
QY 2729 tttaggaatttcaactttagatcatcttggttccacttttatttcaatttatttctg 2788
Db 149589 TTAGGAATTTTAAACCTTAGATCATCTTTCCTCAATTTTATTTTATTTTATTTTGG 149530
QY 2789 ttaatgaacacaggaact-----aatgaacttgaatctgctgttttaagaatttataaaa 2844
Db 149529 TTAATGCAACAGGACTTTAAACAATGAACTTTGATCTCCATTTTAAATATTTCTTTAGA 149470
QY 2845 acattgtgtaactatataatcttgacttgaagacttgaacttcaactacactacagatat 2904
Db 149469 AAATGCAATGTCTATACATACGGGCTGTGAGACTTACCTTCACTACACTACAGGATAT 149410
QY 2905 gatctccatgtatcacaataaactcgcagagatgatttccagagtgctcgatactgtta 2964
Db 149409 GATCTCCACATGATTTCACTAAACCTGCAACAGTATTTTCCAGAGTCTCGATCTATTTTA 149350
QY 2965 attactctcatttagggcgtgaaaaaagtactacagttctgtatatacagctgtgtgct 3024
Db 149349 ATTACACTCTGATTAAGGCTGAAAAAGTATACACCATATTTCTTTATACAGTTGTGTGCT 149290
QY 3025 tttagatgtgttactgttacacagaagtgtgtgactgagac 3067
Db 149289 TTGATGTGTGTACTATATACACAGAGTGTGTGCACTGGGCC 149247

```

RESULT 12  
AC068845/c

LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

AC068845 194407 bp DNA HTG 04-SEP-2000  
Homo sapiens chromosome 19 clone RP11-73711 map 19, WORKING DRAFT  
SEQUENCE. 27 unordered pieces.  
AC068845  
AC068845.3 GI:9966312  
HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 194407)  
Homo sapiens chromosome 19, clone RP11-73711  
Unpublished  
2 (bases 1 to 194407)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barina, N., Bastien, V., Bedalov, F.,  
Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G.,  
Campoliano, A., Castle, A., Choquet, Y., Colangelo, M., Collins, S.,  
Collamore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S.,  
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,  
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,  
Levine, R., Liu, G., Locke, K., Macdonald, P., Marquis, R.,  
McCarthy, M., McEwan, P., McGuck, A., McKernan, K., McPeeters, R.,  
Meldrum, J., Menus, L., Mihova, T., Miranda, C., Mienze, V., Morrow, J.,  
Murphy, T., Maylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, R., Plerre, N.,  
Pisanil, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Testaye, S., Theodore, J., Tirrell, A., Travers, M., Triggiano, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (10-MAY-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 4, 2000 this sequence version replaced gi:8247880.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: MIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L7549  
Center clone name: 737.I.1  
----- Summary Statistics  
Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 181875 bases at least Q40  
Consensus quality: 187842 bases at least Q30  
Consensus quality: 190220 bases at least Q20  
Insert size: 190000; agarose-fp  
Insert size: 191807; sum-of-ctrls  
Quality coverage: 4.7 in Q20 bases; sum-of-ctrls  
Quality coverage: 4.7 in Q20 bases; sum-of-ctrls  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 27 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1755: contig of 1755 bp in length  
\* 1756 1855: gap of 100 bp



RESULT 10  
AF227200 3059 bp mRNA ROD 09-AUG-2000  
LOCUS Rattus norvegicus staufen isoform Stau+16 (Stau) mRNA, complete  
DEFINITION cds, alternatively spliced.  
ACCESSION AF227200  
VERSION AF227200.1 GI:9754865  
KEYWORDS  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 3059)  
AUTHORS Monshausen, M., Putz, U., Rehbein, M., Schweizer, M.,  
Descosseliers, L., Kuhl, D., Richter, D. and Kindler, S.  
TITLE Two rat brain staufen isoforms differentially bind RNA  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3059)  
AUTHORS Monshausen, M. and Kindler, S.  
TITLE Direct Submission  
JOURNAL Submitted (20-JAN-2000) University of Hamburg, Institute for Cell  
Biochemistry and Clinical Neurobiology, Martinistrasse 52, Hamburg  
20246, Germany  
FEATURES  
source Location/Qualifiers  
1..3059  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/issue\_type="brain"  
/dev\_stage="adult"  
1..3059  
/gene="Stau"  
227..1714  
/note="RNA-binding protein; similar to Drosophila staufen;  
contains four RNA-binding domains; rStau+16"  
/product="staufen isoform Stau+16"  
/protein\_id="AAF98119.1"  
/db\_xref="GI:9754866"  
/translation="MKPVDPHSRMOSTVSYMGKGAAYPRPYRPPVPLLYQVELS  
VGGQFNGKMRPPVKHDATARALRTLOSEP,PERLEVNGRESEENLNKSEISQV  
EALRNLVNESEFPLQVARESGPHKKNVTREYSGEYEGEGSKSKSKNA  
RAVLRLPLPAVERVPRPKKSOPTKLOTAPDYGGGNPISRLAOIOQAKKE  
KEPEWMLTERGLPRREFVMQVGHHTAGAGTNKKAKNAENMLETIGFVKPO  
AOPAKPALKSEKTPYKKRGDKRKYTFPPSPDEKSGSNKEDFERPPLSHQQLPAG  
ILPMYEVAAQAVGSGHHTKDFARAAPPAATYATAMTARELLTGISTPTAETILKS  
NLSGHPHGPRTRESEQLYLSRAQGFQVEYKDPKPKNKNCVSLINCSQPLVSH  
GIGKDVESCHDMAALNLIKLSLSDQOSTEMPTGNGPVSAAGRC"  
3043..3048  
polyA\_signal  
BASE COUNT 817 a 778 c 749 g 713 t 2 others  
ORIGIN  
Query Match 39.0%; Score 1253.8; DB 74; Length 3059;  
Best Local Similarity 86.3%; Pred. No. 9,7e-282;  
Matches 1428; Conservative 0; Mismatches 207; Indels 20; Gaps 3;

Db 346 AGTTGAGCTCTCCGTGGAGACAGCAGTTTAAGGAGGAAAGATGAGACACCTGT 405  
563 gaacacgctgctgctcgaagcgttgaagatccctcagaatgagccctgcagagag 622  
Db 406 GAAACATGATGCCACTCCCGAGCCCTGAGGACTCTCAGAGTAGACCCCTTACCAGAG 465  
623 gctggaagtgaatgaaagaaatccgaagaagaatcctaataatctgaataagtc 682  
Db 466 GCTAAGAGTAAATGAGAGAGATCAGAAAGAAACCTTAATTAATCAGAAATTAACCA 525  
683 agtgttgagatgacactaaaggaactgctgtgaattcg----- 726  
Db 526 AGTTTGAAGATTGCGCTGAAGCGAATTTGCGCTGAATTTTGAATTTTCCCTTGAC 585  
727 -aggctgcgggagaggtgagcccccacatgaagaacttttgacagagttcg 784  
Db 586 ACAAGTGGCCCGGAGAGTGCCCGCCACACATGAAGAACCTTTGTGACAGGTTTACT 645  
785 tgggagattctgtyggggaagtgaaaggaagcaagaagatltcaagaagaaatgctgc 844  
Db 646 TGGGGAATTTGTAGGGGAGAGAGAGGAAAGCAAGATCTCCAAAGAAATGCGGC 705  
845 catagctgtctctgagagagctgaaagaaagttaccgcccctgcctgcagtgaacgagtaaa 904  
Db 706 AAGGCGCTGCTGGAGCAGCTAAGAGAGGCTGCCACCTCTACTGCTGTGAGAGCGAGTAA 765  
905 gcttagataaagaagaacaaacccatagtcgaagcagcagagccaggaatattg 964  
Db 766 GCCCAGATCAAGAGAAAGATCAGCCCACTGCAACCTCAAGACTGCCCCGACTATGG 825  
965 ccagagatcaatccagatagtcgagactgcccagatccagcagagcaaaagaagagaaga 1024  
Db 826 CCAAGGATTAATTCCTTTTATGACAGACTGACAGATTCACAGGCAAAAAGGAGAGCA 885  
1025 gccagagatcacgctccctcagacagagagcctccgcgcgcagagagattgtgatgca 1084  
Db 886 GCCGAGTACATCTCTCTTACAGAACAGAGCTTCCACGTCGCGAGGAGTTTGATGACA 945  
1085 ggtgaaggttggaaacacacacctgcagaagaagcggacaaagaagttgcccagag 1144  
Db 946 GGTGAAGTTGGGCAATCACCCGCAAGAGAGGATCAATTAAGATGGCCCAAGCG 1005  
1145 caatgcagcgaagaacatgcttgagatccttggtttcaaaagtcgcgagagagccac 1204  
Db 1006 CAATGCTGCTGAGAGACATGCTGGAGATCTCGGGTTCAAGTCCCGCAGCGACCTGC 1065  
1205 caaacccgacatcaagtcagagagaagacacccataagaacagaggaatggaagaa 1264  
Db 1066 CAAGCCAGCACTCAAAATCAAGAGAGAGACTCCAGTAAAGAAACCAAGAGATGMAAGAA 1125  
1265 agtaaaccttttgaagactgctcctgagagatgaataagagagtagtaataagaagatga 1324  
Db 1126 AGTAACATTTTGTGAACCTTGAACCTTGGGATGAATAAGAACAGTATTAAGAGATGA 1185  
1325 gttcagatgcttattcaatgcatcagcagcgtgcctgctggaattcttccatgctg 1384  
Db 1186 GTTCAGGATGCCCTATCTTATGATGATGAGAGCTGCGAGCGGGATTTCTCCATGATGCC 1245  
1385 cgaagtcgcccagagctgtagagattagtaagaagatcacacaaagaatttaccagagc 1444  
Db 1246 AGAGGTTGGCCAGAGCTGTGCGGGTTAGTCAAGGACACCAATTAAGAGATTTCCCAAGGC 1305  
1445 agtctcgaatcctgcgcaagagcagcagatgcatatgataatgcccagagctgtgtgtag 1504  
Db 1306 AGCTCCAAATCTGCGCAAGGCGCAAGGATGCACTGATATGATGCGGAATTTGTATCGG 1365  
1505 gggcaccctgcgcacagccagagacatlttaagaataaataatccttcaagcagctacc 1564  
Db 1366 GGGCACTGCGCCACAGCCGAGACATTTTAAGAGATTAATCTCTTACAGACATGTAC 1425  
1565 ccatgagcttcaagagagacccctgagcaactgagactatcttccagagtcagagatt 1624



QY	1463	ggccacggtlaactgcatagatagcccgagagtcgttgataggggcaacttcgcccacagc	1522
Db	1306	ggccacggcttaactgcatagatagcccgagagtcgttgataggggcaacttcgcccacagc	1365
QY	1523	cgagacatltttaagaataacatctctcttcagcgcaagtlaccatcagatccttcacagag	1582
Db	1366	cgagacatltttaagaataacatctctcttcagcgcaagtlaccatcagatccttcacagag	1425
QY	1583	accctctagacaactggagctactcttcacagagtcgaagatccagatctcagatatacaaga	1642
Db	1426	accctctagacaactggagctactcttcacagagtcgaagatccagatctcagatatacaaga	1485
QY	1643	cttcocccaaaaacaagaagaacttgatctcttcataatgtctccttcagcaacc	1702
Db	1486	cttcocccaaaaacaagaagaacttgatctcttcataatgtctccttcagcaacc	1545
QY	1703	tcctgatccgcatgtatccgcaagaagatgtgagtcctcgccatgatatgctgcgtcgaa	1762
Db	1546	tcctgatccgcatgtatccgcaagaagatgtgagtcctcgccatgatatgctgcgtcgaa	1605
QY	1763	catcttaaaagtgcgtctgtgagtcggaccaacaagaatcagaagatgccaaagaaggaa	1832
Db	1606	catcttaaaagtgcgtctgtgagtcggaccaacaagaatcagaagatgccaaagaaggaa	1665
QY	1823	cggacaaatgtctgtgtgtgaggtgctgtaaccttcttggccatgaaacattataaaa	1882
Db	1666	cggacaaatgtctgtgtgtgaggtgctgtaaccttcttggccatgaaacattataaaa	1725
QY	1883	tcccaactatatacttgaaaaactcga-aactgtcttgaaaaatttgaaattctgatacc	1941
Db	1726	-cccaactatatacttgaaaaactcga-aactgtcttgaaaaatttgaaattctgatacc	1784
QY	1942	tcccaatggcgccagagaa	1958
Db	1785	tcccaatggcgccagagaa	1801
RESULT	9		
AF061942			
LOCUS	AF061942	2860 bp	mRNA
DEFINITION	Mus musculus RNA-binding protein stauflen mRNA, complete cds.	04-MAR-1999	
ACCESSION	AF061942		
VERSION	AF061942.1	GI:4335944	
KEYWORDS			
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 2860)		
TITLE	Wickham, L., Duchaine, F., Luo, M., Nabl, I. R. and Desgroselliers, L.		
JOURNAL	Mammalian stauflen is a double-stranded-RNA- and tubulin-binding		
MEDLINE	protein which localizes to the rough endoplasmic reticulum		
REFERENCE	Mol. Cell Biol. 19 (3), 2220-2230 (1999)		
AUTHORS	2 (bases 1 to 2860)		
TITLE	Duchaine, F., Luo, M. and Desgroselliers, L.		
JOURNAL	Direct Submission		
MEDLINE	Submitted (28-APR-1998) Biochemistry, University of Montreal, 2900		
REFERENCE	Edouard Montpetit, Montreal, Que H3T 1J4, Canada		
AUTHORS	location/Qualifiers		
TITLE	1. .2860		
JOURNAL	/organism="Mus musculus"		
MEDLINE	/db_xref="taxon:10090"		
REFERENCE	326. .1789		
AUTHORS	/codon_start=-1		
TITLE	/product="RNA-binding protein stauflen"		
JOURNAL	/protein_id="AA017529.1"		
MEDLINE	/db_xref="GI:4335945"		
REFERENCE	/translation="MYRPVPHSMOSTYSYGRGACAPRRYVPPVPLVQVLS		
AUTHORS	VGGQGFNGKMRPVAKHDAFARALRFLQSEPLPERLEVNGREAEENLNKSTISYF		
TITLE	EIALRNLPNVFEVARESGBPMMKNFTRVSGVEFVGEGRGSKRISKNAARAVLEQ		
JOURNAL	LRRPLRPVAVERYKPRIKKKSQETKTAADYGGGMPDISTRADIDQAKKEKEPEYML		





```

Db 1919 ATTGGAGCAAA-CCTGGACAGCCATGCCAGAGAGAG-CCTTGGACGCCGCCAGAGCTAA 1976
Qy 2305 aagcaccagagaataatcaatgtcttccactacagcgtagcacaacttcttagtgagca 2364
Db 1977 AACACACAGAAATCAATGCTTCTACTCAGCGGAGACCACTTTTCTAGTGCCCA 2036
Qy 2365 cggcccaacacccctctgacagtaaccacacatcacaccactgtcttcttccacagtaga 2424
Db 2037 CGGCCCCACACCTCTGAGTACCCACACCATCACACACTGCTTCTTCCCAACAGTGA 2096
Qy 2425 tctgattcttagttcatatttcttctttagatgataagaactatataatttcat 2484
Db 2097 TCGTATCTCTAGTTTCATTTATTTCTTTGGATTGATACACTATATATAATTTTCAT 2156
Qy 2485 ttgaagattctcaatctgatactagtaataaagaagattggaactgtctgagagctg 2544
Db 2157 TTGAGATTTTCTCAATTGATCTAGTTAAATACACAGTTTGGAACCTTGTCGAGACTG 2216
Qy 2545 acttcaataaactcaacacagacatcataatcgtatgtatgtatagacatttt 2604
Db 2217 ACTTATCATATATCTAACCGACAAAGATCATATCATGTGTATGTAGACATTTT 2276
Qy 2605 attctctgactaacccagagacagttcagtgatgcaaatgtgtccctcgttcagc 2664
Db 2277 ATTTTCATTTGACTAACCCAGACAGTTTCACTGATGCATAATTTGTGCCCTCTGTTACG 2336
Qy 2665 tgaacagtcctgagacttcaaaaacacttgaatgaatctccacagttgtataatgga 2724
Db 2337 TGAACAGTCCTGACTTTTCAAAAACCTTGAATGAATCTCCACAGTTGATTAATTTGGA 2396
Qy 2725 caattgaagaattttaaacttagatgatcatattgttccatttatttatttatt 2784
Db 2397 CAATTAGGAATTTTAACTTTAGATGATCATTTGGTTCCATTTTATTTATTTTATTT 2456
Qy 2785 ttgttaagcaaacagagacttaagaacttgcactcgttttaagaattataaaa 2844
Db 2457 TTTGTTAATGCAAAAGAGACTTAAATGAATCTTGTATCTCTGTTTAAAGATTTTAAAAA 2516
Qy 2845 acattgtatcatatcatatgagctcttgagagacttagcttctactacacacagatat 2904
Db 2517 ACATTTGTATCTATATACATATATGCTCTTGAGAGACTTATGCTTACATACATAGATAT 2576
Qy 2905 gatcccatgtagtcataataaacctgacagatgatttccacagatgctcgaactgta 2964
Db 2577 GATCTCAGTACTGATCATATATAACCTGACAGATGATTTTCCAGAGTGCCTGATACTGTTA 2636
Qy 2965 attaatctccatagtaggctgtaaaaagaatgaactaagttctgtatatacagctgttgc 3024
Db 2637 ATTACATCTCCATTTAGGCTGAAAGAAATGACCTACGTTTCTGTATACACAGCTGTGCT 2696
Qy 3025 ttgagctgtgtgtactgtatcacagaagtggtgcactgagagctctgagctgtgtccgta 3084
Db 2697 TTTGATGTTGTGTACTGTATACAGAAAGTGTGTGCACGTAGAGCTCTGCGTGTGTCCGTA 2756
Qy 3085 tggaaaaactgtgtagccctgagatgaatgaactgtcttccatcttctttagcgtggaat 3144
Db 2757 TGGAAAGCCTGGTAGCCCTGCGAGTTAAAGTACGCTTCATTTGTTTACCTGGAAT 2816
Qy 3145 tttctccccaatgaatgtaagtaaaaacttaagtgtttgtcatcaataaagttaact 3204
Db 2817 TTTTCTCCCATGGAATGTAAGTAAAACTTAAGTTTGTTCATCAATTAATGTTAATACT 2876
Qy 3205 aaa 3207
Db 2877 AAA 2879

```

```

VERSION AL133174.15 GI:8573761
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE 1 (bases 1 to 113836)
JOURNAL Direct Submission
Submitted (14-JUN-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT requests: clonerequest@sanger.ac.uk
On Jun 20, 2000 this sequence version replaced gi:8388434.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
RP3-470L14 is from the library RPCT-3 constructed at the Roswell
Park Cancer Institute by the group of Pletier de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
IMPORTANT: This sequence is not the entire insert of clone
RP3-470L14. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true right end of clone RP3-470L14 is at 113836 in this
sequence. The true right end of clone RP1-1556 is at 100 in this
sequence.
FEATURES
source 1..113836
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RP3-470L14"
/clone_lib="RPCT-3"
/misc_feature 3061..3077
/Note="Weak data"
misc_feature 95583..95627
/Note="Single clone region. pUC subcloned to give 12x
coverage. Assembly confirmed by restriction digest data."
BASE COUNT 33974 a 25115 c 24233 g 30514 t
ORIGIN
Query Match 42.2%; Score 1357.4; DB 34; Length 113836;
Best local Similarity 99.9%; Pred. No. 0;
Matches 1369; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
Qy 1838 gtgtggaggtgtgtaaccttcttgccatgaacatataaattccacatatatc 1897
Db 34965 GTGTGGAGGTGCTGAACCTTTCTGCGCATGAAACATTAATAATCCCAACATATATAC 34906
Qy 1898 tgaataactgaacacgtcttgaataatttctgtactctcagtgaggccgagag 1957
Db 34905 TGAATAATACGAAACGCTTTGAAATTTCTGTATACCTCCAGTGGGCCGAGAG 34846
Qy 1958 acacgtgtgtaagatgtgggcagcagcaggaagacacagaacacagaagcgg 2017

```



LYOVELSVGOOFNGKRTROAHNAKALRLIOLNEPLERLVNGRESEENLNK  
SEISOVEFIATLRLPLVNFVAVRSEPHMNFYTKVSEVGEVGEKSKISKNA  
AIAVEELKILPLKPLVNFVAVRSEPHMNFYTKVSEVGEVGEKSKISKNA  
EKEPEHYTLTERGLPRRREPEYMOYVGHNTHEGISTNKVKAKRNAEMLEITGRVP  
QRSHOTRQVGEDETHETKRMKSNLFLNALGEMNTSKNEDFEFVPIYSHOOLPA  
GILPWPVPAOVAVSOGHMKDFLRAAPNPAKATVIMARLELYLGGTSPAEITLK  
NNISGHPHGLPFRSEPOLDYLSRVQGFQVEYDFPKNNNEFEVSLINCSOPPLIS  
HGICKDVSCPHDMAALNITLKLESDOOSTEMPRTGNPMVSCGRC"

BASE COUNT 891 a . 693 c 699 g 783 t

ORIGIN

Query Match 87.0%; Score 2799.4; DB 68; Length 3066;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 2870; Conservative 0; Mismatches 6; Indels 7; Gaps 6;

QY 328 gcataaccctactctgtagaactaaatgcactgtagaacttggaaaaaacaatgt 387  
DB 1 GCATAACCCCTACTGAGAACTAAATGACACTGTCATGAACCTTGAAAAACCAATGT 60  
QY 388 ataaagctgtgtgacctactctgtagatgcaatccaccataactacaatgagagag 447  
DB 61 ATTAAGCTGTGTGACCTTAATCTCGATGCACTATCACTATCAACATGAGAGAG 120  
QY 448 gtgcttatcccccagagtagtacttaccatctccagttccacttacttactaaagt 507  
DB 121 GTGCTTATCCCCGAGAGTACTTTTACCCATTTCACAGTTCACACTTTTATCAAGTGG 180  
QY 508 aactctctgtgagagagacagaacttaatgcaaaagaaagaaagacagctgcgaaac 567  
DB 181 AACTTTCTGTGGAGAGACAGCAATTTAATGCAAAAGAAAGAACAGAGCTGCGAAGC 240  
QY 568 agatagctgtctgcaaaagagttgagagatctctgcaagaatgagccccgcagagagagctg 627  
DB 241 AGGATGCTGTCTCCCAAGGCTGTGAGATCTCTGAGATGAGCCCTGCGCAGAGAGCTGG 300  
QY 628 aggtgaatgaaagagatccgaagaagaatctcaataatctgaaatgaatgaatgaatgt 687  
DB 301 AGGTGAATGAAAGAGATCCGAAGAAATCTCAATTAATCTGAATTAAGTCAAGTGT 360  
QY 688 ttgagattgcaacttaaacggaacttgcctgtgaatttcgaggttgcccgagagagtgcc 747  
DB 361 TTGAGATTGCACTTAAGCGAAGTCTGCTGTGATTTTCAGAGTGGCCCGGAGAGTGGCC 420  
QY 748 caaccacaatgaaagacttctgcaagaagttcggttgaggagagttctgtggggagagtg 807  
DB 421 CACCCCACTGAAGAACTTTGTGACCAAGTTTCGGTTGGGAGTTTGTGGGGGAAGGTG 480  
QY 808 aaggaaagaaagaaagatttcaaaagaaatgcccgaatgctgttcttgaggagctga 867  
DB 481 AAGGAAAGAAAGAAAGATTTCAAAGAAATATGCCCATAGCTGTTCGTGAAGAGCTGA 540  
QY 868 agaatgttaccgccccctgtcagttgaaagagtaaaagctagaatcaaaagaaagaa 927  
DB 541 AGAAGTTTACCGCCCTGCTGCGAGTTGAAACGAGTAAGCTAGAAATCAAAAGAAACAA 600  
QY 928 aaacctatgtaagcaacagacagagccagataatgccaagggatgcaatccgattagcc 987  
DB 601 AACCCATTAATGACACACAGACAAAGCCCAAGATATGCGCAGGGGATCAATCCGATTAGCC 660  
QY 988 gactgcccagatccagcagggcaaaaagagagagagcagagatcaagctcctcaag 1047  
DB 661 GACTGCCCAAGATCCAGACAGGCAAAAAGAGAGAGAGCAGAGTACAGCTCTCTACAG 720  
QY 1048 agcgaagcctcccgcgcgcaagggagttgtgagtgcaagtgaaagtgtaaacacactg 1107  
DB 721 AGCGAGGCTTCCCGCGCCCGCAGAGAGTTGTGATGAGGTGAAGGTTGGAACCAACACTG 780  
QY 1108 cagaagaagcagcacaagaagaagtgccagcagcagcagcagcagcagcagcagcagc 1167  
DB 781 CAAAGGAAGGCGGCAACAAAGAAAGGTGGCCAAAGGCAATGACGCGAAGAAATCTCTGG 840  
QY 1168 agatccttggttcaagtcgccgagcgagcccaaaaccgcagctcaagtcagag 1227

DB 841 AGATCCTTGTTTCAAGTCCGACGCG--CAGCCACCAAAACCCGACACTCAAGTCAGAG 898  
QY 1228 aagaagaaccccaataagaacacagggatgtagaagaagtaacc-tttttgaacctgc 1286  
DB 899 AAGAACACCCCTAAAGAAACAGGAGGAGGAAGAAAGTAACCTTTTGAACCTGGC 958  
QY 1287 tctggggat-gaaaatgagacagtaataaagaagatagttcagagatgacctactaag 1345  
DB 959 TCTGGGATGGAATAATGGAGCTAGTAATTAAGAGATGAGTTGAGATGCTTATCTAAG 1018  
QY 1346 tcatagcagctgtcctgtcgtgaattcttccatgggtgcgccagtcgccccagctag 1405  
DB 1019 TCATCAGCAGCTGCGCTGCTGAATTCCTCCATGTGGCCGAGTGCAGGCTGTAG 1078  
QY 1406 agttgtagaagacatcacacaaagaatttaccagggcagctccgatactgcgaagc 1465  
DB 1079 AGTTAGTCAAGGACATCACACAAAGATTTCACAGGGGAGCTCGAATCTGCCAAGGC 1138  
QY 1466 caagtaactgcatgtagcccgagagttgtgtataggggacactgcgccacagcga 1525  
DB 1139 CAGGTAAGTGCATGATAGCCGAGAGTTGTGTATGGGGACCTGCCACAGCCGA 1198  
QY 1526 gacattttaagaataacatctcttcagggccaagtlaccccaaggacctctcaagagac 1585  
DB 1199 GACCAATTTAAAGAAATTAACATCTCTTCAGGCGACGATCCATGAGCCTCAGAGAC 1258  
QY 1586 ctcgtagcaactgtagtactcttccagagtcagagagatccaggttgatacaagaact 1645  
DB 1259 CTTGTGCAACTGAGCATCTTTCAGAGTCCAGGAGATTCCAGGTGAATTAAGACTT 1318  
QY 1646 ccccaaaaacaaagaagaattgtatctcttcaatgtcctctcagcacctct 1705  
DB 1319 CCCCAAAACAAAGAAAGAAATTTGTATCTCTTATCAATTCCTCTCAGCCACTCT 1378  
QY 1706 gatcagccaatgtatcogaaagatgtgagtcctgcagcaatgatalgctgtgcgtgaac 1765  
DB 1379 GATCAGCCATGTGATGCGCAAGAGATGTGAGTCTGATCAATTCCTCTCAGCCACTCT 1438  
QY 1766 cttaaagttgtcgtgtagtgtgacccaacaagaatagatgcaagaagaagaagag 1825  
DB 1439 CTTAAAGTTGCTGTGATGATTTGACCAACAAAGTACAGAGATCCCAAGCAAGAAACGG 1498  
QY 1826 accaatgtctgtgtgagaggtgtcgaaccttcttgccatgaaacatlaataatcc 1885  
DB 1499 ACCAATGTCTGTGTGGAGGTGCTGAACCTTTCTGCGCAAGAACATTAATAATCC 1558  
QY 1886 caacatataactgaaataactgaaactctctgaaatctgaaatcttctgaatacttca 1945  
DB 1559 CAACATATTAAGTAATAATTAAGTAATACTCTTTGAATAATTTGGAATTTCTGATACCTCA 1618  
QY 1946 gtggcgcgagagacacggtgtggaagaatgtgtggcagcagcagaggaagaagaagaac 2005  
DB 1619 GTGGCGCGAGACACAGGGGTAAAGATGTGGCGCAGCAGGAGGGAACAAACAGAAAC 1678  
QY 2006 aaaaagagcgcgctgtggcc-ggctgagactgtgtgagggttggtgtgtgtgtgtgtgt 2064  
DB 1679 ACAAGAGGCGGCTGTGGCGGGCTGTGAGTGTGGGGGTGTGTGTGTGTGTGTGTGTGT 1738  
QY 2065 tgaactgtgcgtccctacgcaatagcagctgtcgtgtggaagaagaaggtgtgccagccag 2124  
DB 1739 TGACGTGGCGGTGCTTACCAATAGCAGCTGCTGTGGGAAGAGGGGTGCCACAGCCAG 1798  
QY 2125 ctgtgtctcccgagacacagcagatccacacccctggggacactcgtgtgtgtgttttt 2184  
DB 1799 CTGTGTCTCCCGGACACACAGATCCACACCTGGGACCTGCTGTGTGTGTGTGTGTGT 1858  
QY 2185 ttcccccgtgtgaaagaagaagaagcagacccccttccaaagctgtgctcacctcagacac 2244  
DB 1859 TTTCCCTGTGTGAAGAAAGAAAGGCGACGACCCCTTCTCAAGCTGTGCTCACTGACAGC 1918  
QY 2245 attgagacaacccctggacagcgaatgccaagagagagccttgaacggcccaagactaa 2304

QY 1632 gaalacaaagactccccaacaaacaaagacgaatttctatcttcaattgtctcc 1691  
 |||||  
 Db 1921 GAATACAAAGACTTCCCAAAACAAACAGAGAAATTTATCTTATCAATTCCTCC 1980  
 QY 1692 tctcagccactctgtatcagccatgtglatcgcgaaggaatgtgaagtcctgcataatg 1751  
 |||||  
 Db 1981 TCTCAGCCACTGTGATCACCATGTGTGCGCAAGATGTGGAGTCTCCCTCCATGATATG 2040  
 QY 1752 gctgcgctgtaacacttaaaagtctgtctgtatgtgaacaaagtaagaatgtcca 1811  
 |||||  
 Db 2041 GCTGCGCTGAACATCTTAAGTTGCTGTCTGAGTTGGACCAACAAATACAGAGATGCCA 2100  
 QY 1812 agaacagaaacggaacacatgtctgtgtgtgagagtgctgcgaaccccttctgcgcataa 1871  
 |||||  
 Db 2101 AGAACAGGAAGACGACCAATGTCTGTGTGGAGGTGCTGAACCTTTTGTGGCATGA 2160  
 QY 1872 ccaatlataaaatcccaacataatactgaataactgtaaacctgtcttgaataattggaat 1931  
 |||||  
 Db 2161 CCATTTATTAATCCCAACATATATGTAAGAAATACGAAACCTCTTGAATAATTTGGAA 2220  
 QY 1932 ttctgtacctccagtggtgcgcgaagagaacagtggtgtaagagaatgtggaagagcagg 1991  
 |||||  
 Db 2221 TTCTGATACCTCCAGTGCGCGAGAGACAGGTGGTAAAGAGATGTGGCAGCAGCAGG 2280  
 QY 1992 aagacaacagaaacacaaagagagcgctgtgtgcgcgcgtgagactgtgtgtgtgtgtgt 2051  
 |||||  
 Db 2281 AAGACAACAAACAAACAGAGAGCGGCTGTGGCGCGCTGGACCTGTGTGGGTTGTGTGT 2340  
 QY 2052 gatgccaactcgtgtgacctgtgcgttccctacgaatagcagctgtcctgtgtgtgtgtgt 2111  
 |||||  
 Db 2341 GATGCGCACACTGGGTGACCTGTGGCGCTCCCTACGCAATAGCAGCTGTGTGGCAAGACG 2400  
 QY 2112 gctgcgcgaacagctgtgtctccgcgaacacagacagatccacacccctgtgcacccctcgt 2171  
 |||||  
 Db 2401 GCTGCCAGCAGCTGTGTCTCCGGGACACAGCAATCCACACCTGTGGGACCTCCGT 2460  
 QY 2172 gtttggctcttttttcccccgtgtgtgaagaagaacgacgacgacccctctcaagctgtg 2231  
 |||||  
 Db 2461 GTTTGGCTTTTTCCTTCCCTGTGTGAAGAGAGAAAGGAGCAGCCCTCTCAAGCTGG 2520  
 QY 2232 ctcaactcaagacacatgtggagaacacccctgtgacagcagctgtgcagagagagccttgcag 2291  
 |||||  
 Db 2521 CTCACACAGACATTTGGGAGCAAAACCTGTGACAGCCATGCAAGAGGCGCTTTTGACCG 2580  
 QY 2292 gccccaagagctaaagacacacagagaatacaaatgtcttctactagcgagacccaactt 2351  
 |||||  
 Db 2581 GCCCCAGAGCTTAAGACACAGAGAAATATGCTCTCTACTCAGCGTGACCCAACTT 2640  
 QY 2352 ttctagtggtgcacagcgcacacacacactctctcagtaaccacacacacacacacacacacac 2411  
 |||||  
 Db 2641 TTCTAGTGTCACAGGCGCCACACACCTCTCGAGTACCCACACACACACACACACACACAC 2700  
 QY 2412 cttcacaagtgatctgtatcttctagtttcaatacttcttcttctgattgataagacactat 2471  
 |||||  
 Db 2701 CTTCACAAACATGTATCTGTATCTTAGTTTCATATTTCTTTGATTTGATATGACACTAT 2760  
 QY 2472 ataaatttcatctttagaatttctcaatgtatctagtttaataagacaggtttgaaac 2531  
 |||||  
 Db 2761 ATAAATTTTCATTGTGAATTTCTCAATGTATGTAGTTAAATAGACAGTTTGGAAAC 2820  
 QY 2532 ttgtctgaagctgaacttcaataatactaacgacaagaatcataatcctgtatgtg 2591  
 |||||  
 Db 2821 TTGCTGAGAGCTGACTTTATCATATCTATACGACAAAGATATATCCATGTGTATGTG 2880  
 QY 2592 gttagaacatttattatctatctatctgaacacagagacgtttcagtgatgcaaatgtgtgc 2651  
 |||||  
 Db 2881 GTTAGACATTTTATTTATTTATTTGACTTAACCCAGACAGTTTCAGTGAATGCAATTTGTGTC 2940  
 QY 2652 cctctgttcaagctgaacagctcctgtgacttcaaaaaaccttgaataagcttccacagct 2711  
 |||||  
 Db 2941 CCTCTGTGTTACGCTGAAGACGCTCGACTTTTCAAAAAACCTTTGAATTAAGTCTCCACAGT 3000

QY 2712 tgataaattggaacaaatttagaattttaaactttagatgatalcatattggttccatttta 2771  
 |||||  
 Db 3001 TGATATAATTGGACAATTTAGTAATTTAAACCTTAGATGATCATTTGTTCATTTTAA 3060  
 QY 2772 ttcaatttatttatttgtttaaagcaaacagactttaaataagacttgaacttgaacttctgttttaa 2831  
 |||||  
 Db 3061 TTTTCATTTTATTTTGTGTATATGCAACACAGACTTAATATGACCTTTGATCTGTGTTTAA 3120  
 QY 2832 agattattaaaaaacatttggatctatcataatagcttctgtgaagacttgaacttcaacta 2891  
 |||||  
 Db 3121 AGATTTATTAATAAACATTTGTGTATCTATACATATGCTCTTGAGAGACTTACCTTCACTA 3180  
 QY 2892 cactacagaataltgatalccatgtatglatcataataaaccttcgaagtgatlttccaagatg 2951  
 |||||  
 Db 3181 CACTACAGATATGATGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3240  
 QY 2952 ctcgataactgttaataatcatctccatctgaagcctggaagaaagaaatgaactgaagtttctgata 3011  
 |||||  
 Db 3241 CTGGAATCTGTTAATTAATCACTCCATTTAGGGCTGAAAGAAATGACCTACGTTCTGTATA 3300  
 QY 3012 cagctgtgtctcttgaatgtgtgttactgtacacagaagtggtgtgacgtgaagcctgtg 3071  
 |||||  
 Db 3301 CAGCTGTGTTGCTTTTGTATGATTTGTGTACTGTACACAGAAAGTGTGTGACTGAGGCTGTG 3360  
 QY 3072 cgtgtgtccgtatggaacacactgtgtagccctgcaggttaagtaactgtcttccatctcatg 3131  
 |||||  
 Db 3361 CGTGTGCTCGTATGGAACACCTGTGAGCCCTCGAGACTTAAGTACTGCTTCATTTGATTTG 3420  
 QY 3132 tttaagcttgaatttcttcccaatggaatgtgaatgaataaactgaagtgttctgttcaat 3191  
 |||||  
 Db 3421 TTTACGCTGGAAATTTTCTCCCATGGAATGTAAACCTTAAGTGTGTGATCATCAAT 3480  
 QY 3192 aaatgttaacttcaaaaaa 3217  
 |||||  
 Db 3481 AAATGTAATACTTAATAAAAAA 3506

RESULT 6  
 LOCUS HSA132258 3066 bp mRNA PRI 14-MAY-1999  
 DEFINITION Homo sapiens mRNA for stauufen protein, partial.  
 ACCESSION AJ132258  
 VERSION AJ132258.1 GI:4572587  
 KEYWORDS stauufen gene; stauufen protein.  
 SOURCE human.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 3066)  
 AUTHORS Falcon A.M., Fortes P., Marion R.M., Beloso A. and Ortin J.  
 TITLE Interaction of Influenza virus NS1 protein and the human homologue  
 of Stauufen in vivo and in vitro  
 JOURNAL Nucleic Acids Res. 27 (11), 2241-2247 (1999)  
 MEDLINE 99263025  
 REFERENCE 2 (bases 1 to 3066)  
 AUTHORS Ortin J.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-JAN-1999) Ortin J., Molecular and Cellular Biology,  
 Centro Nacional de Biotecnologia, Cantoblanco, 28049 Madrid, Spain  
 FEATURES  
 source  
 1..3066  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 36..1526  
 /gene="stauufen"  
 36..1526  
 /gene="stauufen"  
 /function="gRNA binding protein"  
 /codon\_start=1  
 /product="stauufen protein"  
 /protein\_id="CAB40082.1"  
 /db\_xref="GI:457258"  
 /translation="MKIGKKPMYKVPDPSRMQSTYNYNMRGAVPPRYFPVPVPL

/translation="MSQVOVONPMSALSGSQLINKNOSLIQPLMSIPSTSLPS  
ENAGPIONSALPSASITSTSAASISITPEVLAINKLKPKPKVKKVDPVSRMST  
LYNMRGASVPPRYPPVPPVPLLOVELSVGGGQFNKGRKROAKRDAAKARIL  
ONEPLPERLVNGBRESSEENLINKSISVFEALAKRNPUNPEVAREGPPRMKPEVT  
KVSQGEYVGBGSKKISKNAATIAVEIKLPLPLAVEVAVRIRIKTKPIVKPO  
TSPEVGQGINPISRLAQIOAKKEREPEYTLITTEGLPRRREVVQVGVNTAGTG  
TNKKVAKENAEEMLEIIGFYVPOROPKPKALKEEKPPIKPGGKRVTEPEPSGD  
ENGTNKEDEFEEMPYLSHOOLPAGILPMVPEVAGVAVGSGHHTDFTRAAPNPAKAT  
VTAMARELYIGSTPPTAETILKNKISSGHDHGLTPRSELDQVLSRVQGOVREYKD  
EPKNNKNEFVSLINSSOPPLISHIGIDVESCHMAALNLIKLSLDOOSTEMPR  
GNGPMSVCGRC

BASE COUNT 965 a 875 c 790 g 876 t  
ORIGIN

Query Match 90.6%; Score 2914.8; DB 10; Length 3506;  
Best Local Similarity 91.7%; Pred. No. 0;  
Matches 3215; Conservative 0; Mismatches 2; Indels 289; Gaps 1;

QY 1 actctctgcggagctgcggagccctgaagcgtcttcagcgcttggcgcgcgctgcgt 60  
DB 1 ACTTCTGCGGGGCTCGGGGCGCTTACGCGTTTGCGGCGGCGCTGCGCGT 60  
QY 61 ctctctgcgctccgcgtctcttgacgcgtcccccgcggccgcggcgccgcctc 120  
DB 61 CTTCTCGGCTCCGCTCTTGTGACGCGCTCCCGCGCGCGCGCGCGCTC 120  
QY 121 ctccagcgcaactcgctcctctccctcctcctcctcctcctcctcctcct 180  
DB 121 CTTCCTTCCCTCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180  
QY 181 ctctctccctcctcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 240  
DB 181 CTTCCTTCCCTCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
QY 241 agcagccagagttatlaaaccacttaacctlcagaacttgaacaagaacaatltg 300  
DB 241 AGAGCCAGAGTTATTAACTTAACCTCTCAGAACTGAACAAAGACAACTGTCC 300  
QY 301 tggaaagccctcttlttaaaaa----- 322  
DB 301 TGGAAAGCCCTCTTTTAAAAAAGGTAGAACTTACACTTACATTAACCTG 360  
QY 323 ----- 322  
DB 323 ----- 322  
QY 361 CACTGAACCTGTTTACCTGCACTTTGTTCACTTTGTTGAAGTGCATGTCTCAAGTT 420  
DB 361 CACTGAACCTGTTTACCTGCACTTTGTTCACTTTGTTGAAGTGCATGTCTCAAGTT 420  
QY 323 ----- 322  
DB 323 ----- 322  
QY 421 CAAGTCAGTTGAGAACCAATCTGCTCTCAGGGAGCCAAATFACGAAAGAAC 480  
DB 421 CAAGTCAGTTGAGAACCAATCTGCTCTCAGGGAGCCAAATFACGAAAGAAC 480  
QY 323 ----- 322  
DB 323 ----- 322  
QY 481 CAGTCTTCTCTCACAGCCTTTGATGAGTATCTTACTACTAGCTCTGCGCTCT 540  
DB 481 CAGTCTTCTCTCACAGCCTTTGATGAGTATCTTACTACTAGCTCTGCGCTCT 540  
QY 323 ----- 322  
DB 323 ----- 322  
QY 541 GAAATGACAGTAGACCACTTCAAACTCTGCTTACCTCTGACATCTATFACATCCACC 600  
DB 541 GAAATGACAGTAGACCACTTCAAACTCTGCTTACCTCTGACATCTATFACATCCACC 600  
QY 323 ----- 322  
DB 323 ----- 322  
QY 601 AGTCGACGTGCAGAAAGCAATACCCCTACTGTAGAACTAAATGCATGTGCATGAACATT 660  
DB 601 AGTCGACGTGCAGAAAGCAATACCCCTACTGTAGAACTAAATGCATGTGCATGAACATT 660  
QY 372 ggaaaaaaaccaatgtataagcctgttgacccttactcgcgaatgaagtcacacta 431  
DB 372 GGAATAAACCAATGTATAAGCCTGTGACCTTACTCTGAGTGAAGTGCACCTATFAC 720  
QY 432 taacaatagagagagtgcttatcccccgaggtacttttaaccatttccagttcaact 491  
DB 432 TACAATAGAGAGAGTGCTTATCCCGAGGTACTTTTACCACTTTCCAGTCT 780  
QY 492 ttacttatacaatggaaactctgtggagagcaacaatttaagcaagaagaagaca 551  
DB 492 TTACTTATCAAGTGAACCTTCTGTGGAGAGACAGCAATTTAATGCCAAGAAAGACA 840

DB 781 TTACTTTATCAAGTGAACCTTCTGTGGAGAGACAGCAATTTAATGCCAAGAAAGACA 840  
QY 552 aaacagcgctgcgaacacacatctcgtgcgaagaggttgaagatcctgaaatgaagccc 611  
DB 552 AAGCAGCGCTGCAGAACACATCTCTGTCGAAGCCTTGAAGATCTGCAGATGAGGCC 900  
QY 612 cgcgcgaagaagctggaagtgaatggaagaagatccgaagaagaatltcaataatc 671  
DB 612 CCGCCGAGAGAGCTGAGGTGAATGGAAGATTCGAGAAAGAAATTCATATAATCT 960  
QY 672 gaaataagtcgaagtgttgagaltgcacttaacggaacttgcctgttgaaattcgagtg 731  
DB 672 GAAATAAGTCAAGTGTGTGAGATTGCACTTAAACGGAACCTGCTGTGATTTCCAGGTG 1020  
QY 732 gcccggaagatggtgccaccacacatgaagaacttgttacaaggtttcgttggagg 791  
DB 732 GCCCGGAGAGTGGCCACCCACATGAAGAACTTGTATACAAAGTTTGGTGGGAG 1080  
QY 792 ttgttgagggaaggtlaaaggaaagcaagaagatttcaagaagaaatgcgcacatagct 851  
DB 792 TTTGTGGGGGAAGGTGAAGGAAGCAAGCAAGATTTCAGAAATAATCCCGCATAGCT 1140  
QY 852 gtctctgaagagctgaagaagtlaccgcctcgtcctgcagtgtgaacgagtaagcctaga 911  
DB 852 GTTCTTGAGAGAGCTGAAGAAGTTACCGCCCTGCTGCACTTGAACGAGTAAAGCTTGA 1200  
QY 912 atcaaaagaagaaacaaacccatagtcgaagccacgaagcccgaaatattgccaaggg 971  
DB 912 ATCAAAAGAAACAAACCAACCATAGTCAAGCCACACAAAGCCCAATATGCGCGAGG 1260  
QY 972 atcaatccgattagccagatggtccagatccagcaagcaaaagaagaagagccagag 1031  
DB 972 ATCAATCCGATTAGCCGAGTGTGCGCGATCCAGATCCAGCAAGCAAAAGAGAGGCGAGAG 1320  
QY 1032 taacagctccctcaacagagcgaagcctccgcgcgagggagtttgtatgaatgaag 1091  
DB 1032 TACACCTCTCTCACAGAGGAGGCTCCCGCGCGAGGAGTTGTGATGCAAGGTGAAG 1380  
QY 1092 gtggaaacacacatgcgaagaagagggcagcaaaagaagtggaagcgcaatgta 1151  
DB 1092 GTTGGAAACACACCTGCAGAAAGAGGAGCGGACCAACAAAGGTGSCAAAGCCCAATGCA 1440  
QY 1152 gccgaagaacatgctgagatcctctgttccaagtccgaagcgagcccaaaccc 1211  
DB 1152 GCCGAGAACATGCTGAGATCTTGTGTTCAAAAGTCCCGCAGCGGAGCCCAACCAACC 1500  
QY 1212 gcaatcaagtcgaaggaagaagacccataaagaacccaagggatggaagaaatgaacc 1271  
DB 1212 GCACCTAAGTCAGAGAGAGAAACACCATTAAGAAACCAAGGGATGGAAGAAAGTAAACC 1560  
QY 1272 ttttttgaacctggtctgaggatgaaatggagactagtaataaagaagatgaagttcag 1331  
DB 1272 TTTTGTGAACCTGGCTCTGGGATGGAATGGACATGAATAAAGATGAGTTCAGG 1620  
QY 1332 atgacctatcgaatcaagtaagcagctgctgtggaattcttccatgtgtgcgcgaagtc 1391  
DB 1332 ATGCTTATCTAAGTATATAGAGCTGCTGCTGGAATTTCTTCCATGATGCTGCGGAGTTC 1680  
QY 1392 gcccaagctgtagagtttagtcaagagacatcaacccaagaatttlaaccagggcagctcg 1451  
DB 1392 GCCCAGGCTGTAGAGATTAGTCAAGSACATCACACCAAGATTTTACAGGGAGAGTCCG 1740  
QY 1452 aatctgcgaagcgaagcgaatgcatgatatagccgaagagttgtgtatgggggagcc 1511  
DB 1452 AATCTGCGCAAGCGCACGCTTAAGTGCATGATAGCCGAGAGTGTGTGATGAGGGGAGCC 1800  
QY 1512 tggccacagcggagacatttaagaataacatctctcgaagccagttacccaatga 1571  
DB 1512 TGGCCACAGCGGAGACATTTTAAAGAAATTAACATCTTTCAGGCCACCTACCCATGGA 1860  
QY 1572 cctctcaagagacccctctagcaactgactatcttccagagttccagagattccaggt 1631  
DB 1572 CCTCTCAGAGACCCCTCTAGCAACTGGAATCTTCTTCCAGAGTCCAGGAGATTCCAGGTT 1920



TITLE	Direct Submission
JOURNAL	Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900 Edouard Montpetit, Montreal, Quebec H3F 1V4, Canada
FEATURES	Location/Qualifiers
SOURCE	1..3142
gene	/organism="Homo sapiens"
CDS	/db_xref="taxon:9606"
	/chromosome="20"
	/map="20q13.1"
	1..3142
	/gene="STAU"
	288..1778
	/gene="STAU"
	/function="RNA binding protein"
	/note="alternatively spliced transcript"
	/codon_start=1
	/product="stau protein"
	/protein_id="AADI7532.1"
	/db_xref="GI:4335951"
	/translation="MKLGKSKPMYRPVDYSMSQSYNNMKGAVPERFYFPVPVPL LYQVELSVGGQENFGKGTROAKHDAALKALRIIONEPLDERLEVNRESEELNKL SEIQVEIAGLKRNLPNVEFARESGPPHMKNPTKYSGEFGEGESKISKNNK AIATLEELKKLPPIPAVERVPRITKKTKPIVKPOTSPERGOGINPSRLAOIAKK EKPEPYTLIERGLPIRRREVPMOYKVGHNHRAEGTGNKKVAKRNAEMLEILGKVY ORQTPFKALSSEETPKIRPGDKRVTFEEESDENSTSKDEFRMPYLSHOOLPP GLIMPEVPAOVAVSQGHHTKDPTRAIPNAKATVTAMIARELLYGTSPTAEFIL NMISGVHPHGPLTRPSEOLDYLSRVGFQVEYEDFPNNNEFVSLINCSSOPPLIS HGIGDYESCNDMALNLTIKLISELDOSTEMPRTGNGMVCGRC"
BASE COUNT	862 a 778 c 733 g 769 t
ORIGIN	
Query Match	94.9%; Score 3053.8; DB 10; Length 3142;
Best Local Similarity	97.6%; Pred. No. 0;
Matches 3140; Conservative	0; Mismatches 2; Indels 75; Gaps 1
Oy	1 actctcgcgaggtcgtagcgcttgtagcgtcttcacgclttgcgcggctgcgagt 60
Db	1 ACTTCCTGCCGGGCTGCGGGCGCTTGACGCGCTTACAGCTTTGGCGGCGCTGCGGCT 60
Oy	61 ctctctggctcccgcttcttcattgaaccgctcccccccgggccggcgccgcgcctc 120
Db	61 CTCTCTGGGCTCCCGCTTCCTTTGACGCGCTCCCCCGGCGCGGCGCGCGCCTC 120
Oy	121 ctcaaggcgcaactcgacctcttcctccctcgctccctctctcctctccctttttcct 180
Db	121 CTCGACGGGCACTCCGCGCTTCCCTCTCCCTTGCTCCCTTCTTCCTTTTTCCTT 180
Oy	181 ctctctccctctcctcgccgcgcccacgagcgccgagcgccggcggggagcagctcgaa 240
Db	181 CTTCTCTCCCTCTCTCGCGCGCACGCCAGAGACGCGCGGGGAGACGATCGGAGC 240
Oy	241 agcagcgagagttatttaaccactaactcagaactgaaacaaaagaacaatgltcc 300
Db	241 AGCAGCC----- 247
Oy	301 tggaagcgccctcttttaaanaagaagacataaccctactgtagaacataatgacagt 360
Db	248 -----AGAAAGCATTAACCCTACTGTAGAACATAATGCATGT 285
Oy	361 gcatgaaccttggaaaaaaaaaaaccaatgataagcctgtgtgaaccttactctcgatgacgt 420
Db	286 GCATGAACCTTGAAAAAAAACAATGTATAACCTGTGACCTTACTCTCGGATGCAGT 345
Oy	421 caactataactaacaacatgagaagagtgcttatctcccgagggtactctttaccatttc 480
Db	346 CCACCTATTATACATCAACATGAGAGGAGGTATTCGCCCGGAGGTACTTTTACCCATTTTC 405
Oy	481 caattcacacttactttaagaigtgaactctctgtggaagacagcaatttaagcga 540
Db	406 CAGTTCACACTTTACTTTATCAAGTGGACACTTCTGTGGGAGAGACGACATTTAATGGCA 465
Oy	541 aaggaagaagaagacagctgcgaaacacgatgctgctgcgaagaagcgttgtgaagatctcgc 600

Db	466	AAGGAAAGACAGACAGCGTCCGAACACAGTCCTCTCCCAACACGTTGGAGATCTGC	525
Qy	601	agaatgagcccttcgcaagaagctctgaggtgaaatgaaagaatccgaaagaanaatc	660
Db	526	AGATGAGCCCCCTCCCGAGAGAGGCTGGAGTGAATGGAAGAAATCCGAAGAAGAAATC	585
Qy	661	tcaataaacttgaaataagtcgaagtgtgttgaaatgtgacttaacggaactgtcgttga	720
Db	586	TCAATAAATCTGAATATGAATAGTCAAGTGTGGATTGCACTTAAACGGAATCTGCTGTGA	645
Qy	721	attccgaggttgcccgaggaagtgtgcccaccaccatgaaagactttgtgcccgaagttt	780
Db	646	ATTTCGAGGTGGCCCCGGGAAAGTGGCCACCCCACTGAAGAACTTTGTGACCAAGGTTT	705
Qy	781	cgatttgaggaatttctgtgaggaagaagtgaagggaaagaagaagatttcaagaanaatg	840
Db	706	CGGTTGGGGAGCTTTGTGGGGGAAGGTGAAGGAAAAAGCAAGATTTCAAGAAAAATG	765
Qy	841	ccgcacataagctgtctcttgaaagagctgaagaagttacccgcctgtcctgagttgaagag	900
Db	766	CCGCATAGCTGTCTTGAAGAGCTGAAGATGTACCGCCCTGCTGAGTTGAACAG	825
Qy	901	taagccttagaaatcaaaaaagaacaaaccatagttcaagccacagagcaagccagaat	960
Db	826	TAAACCTTAGAATCAAAAAGAAAAACAACCCCTAATGTCAGGACACAGAACCCAGAT	885
Qy	961	atggccagggatcaatccgattagccagctgtgcccagatccagcaaggaacaanaagaga	1020
Db	886	ATGGCCAGGGGATCAATTCCTGTTAGCCGACTGGCCCAAGATCCAGCAGGCAAAAAAGAGA	945
Qy	1021	aggagccagagtagacgctctccacagagcgagcctcccgcgccgaggaagtttga	1080
Db	946	AGGACCAGAGTATACACGCTCCTCAACAGCGAGGCTCCCGCGCGAGGGAGTTTGGA	10050
Qy	1081	tgcaggtgaaggttctggaaccacaactctgcaagaagcaaggcccaagaaggtgtggca	11400
Db	1006	TGCAGGTAAAGTTTGGGAACCACTGCACAAAGAAAGGAGGCAACAAGAGTGGCCA	10655
Qy	1141	agcgcaatgcagccgagaacatctgtgagatctcttggttcaaaagtcgccgagcgagc	12000
Db	1066	AGCGCATATGCAAGCCGAGAACATGCTGGAGATCTTGGTTTCAAAATGCCGACGGCAGC	11225
Qy	1201	ccaccaaaaccgcaactcaagtctagaggagaagaacaccataaagaacccaaggaatgaa	12600
Db	1126	CCACCAAACCCGCACTCAAGTCAGAGAGAAGAACCCCTAAAGAAACCAAGGGATGGAA	11855
Qy	1261	gaaaagttaaccttttgaacctgtgtctgtgggaaatgaataatgtgagtaataaagg	13200
Db	1186	GAAAAGTAACCTTTTGTGAACTGTGCTGTGGGATGAAATGGGACTACTAATTAAGAGG	12450
Qy	1321	atgagttcaagatgtcctatcttaagtaatacagcagctgtgtgtgaactcttccacatg	13800
Db	1246	ATGAGTTTAGAGATGCTTATCTAATAGTATATAGCAGCTGCTGCTGGAATTTCTCCCATG	13050
Qy	1381	tgcggaggtgcgccagagctgttagagttatgtaagaacatacaccaagaatttaccata	14400
Db	1306	TCGCCGAGGTGCCCGAGGCTTAGAGATTAGTCAAGCAATCACACCAAGATTTTACCA	13650
Qy	1441	ggagcagctcgaatctctgtgccaaagccaggttaactgcataagccggaagttgttgt	15000
Db	1366	GGCGAGCTCCGAATCTGTGCCAAGGCCACGGTAACTGCAAGTAAGCCGACAGAGTTGTG	14225
Qy	1501	atgagggaagcttgcgcccaacgagagacatttaagaataacatcttccaaagcag	15600
Db	1426	ATGGGGGACCTCGGCCCAACGACCGAGCACTTTTAAAGAAATTAACATCTTTACAGGCACG	14850
Qy	1561	taccocatggaactctcaagagacccctctgagcaacttgaactcttccagatccag	16200
Db	1486	TACCCATGGAACCTCTACGAGACCTCTGTAGCAACTGGACTATCTTTCCAAAGTCCAGG	15450
Qy	1621	gattccaggttgaatacaaaagacttcccacaaaacaagaagaaatttgaatcttcta	16800

QY 1369 tctctcccaatgctgcccgaagctcgcccaggtctgtagagtttagtcaagaacatcaccca 1428  
 |||||  
 Db 1501 ttcttcccaatgctgcccgaagctcgcccaggtctgtagagtttagtcaagaacatcaccca 1560  
 QY 1429 aagaattttacaaagggcagctccgaatccctgccaagggcgaagctgaactgcataagccc 1488  
 |||||  
 Db 1561 aaaaattttacaaagggcagctccgaatccctgccaagggcgaagctgaactgcataagccc 1520  
 QY 1489 gaagaattttgtagttagggggaactcgcccacagccgaagaccattttaagaataacatct 1548  
 |||||  
 Db 1621 gaagaattttgtagttagggggaactcgcccacagccgaagaccattttaagaataacatct 1680  
 QY 1549 ctctcaagccacgtaccccaatggaactctcagaagaccctctgagcaactgagatctctt 1608  
 |||||  
 Db 1681 ctctcaagccacgtaccccaatggaactctcagaagaccctctgagcaactgagatctctt 1740  
 QY 1609 ccgaagctccagaagatctccaggtctggaatacaagaactctcccaaaaacaaagaagaat 1668  
 |||||  
 Db 1741 ccgaagctccagaagatctccaggtctggaatacaagaactctcccaaaaacaaagaagaat 1800  
 QY 1669 tctatctctcaatcgaattgtctctcgaagccactctgatacgaagccatgtagtgcgaag 1728  
 |||||  
 Db 1801 ttcttattctcttattcaattgctctctcgaagccactctgatacgaagccatgtagtgcgaag 1860  
 QY 1729 atgtggaagctctgccaatgataatgctgctgagacatcttaagctgctgctgagatctg 1788  
 |||||  
 Db 1861 atgtggaagctctgccaatgataatgctgctgagacatcttaagctgctgctgagatctg 1920  
 QY 1789 accaaacaagaatgaagaatgccaagaagaagaagaagaagaagaagaagaagaagaagaat 1848  
 |||||  
 Db 1921 accaaacaagaatgaagaatgccaagaagaagaagaagaagaagaagaagaagaagaagaat 1980  
 QY 1849 gctgaacactctctgccaatgataatgccaagaagaagaagaagaagaagaagaagaagaat 1908  
 |||||  
 Db 1981 gctgaacactctctgccaatgataatgccaagaagaagaagaagaagaagaagaagaagaat 2040  
 QY 1909 aaactgctctgaagaatctggaatctctgatacctcagctgagcgaagaagaagaagaat 1968  
 |||||  
 Db 2041 aaactgctctgaagaatctggaatctctgatacctcagctgagcgaagaagaagaagaat 2100  
 QY 1969 aaagaatctgagcagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaat 2028  
 |||||  
 Db 2101 aaagaatctgagcagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaat 2160  
 QY 2029 tgaagctgctgagcgt 2088  
 |||||  
 Db 2161 tgaagctgctgagcgt 2220  
 QY 2089 gcaagctgctgagcgt 2148  
 |||||  
 Db 2221 gcaagctgctgagcgt 2280  
 QY 2149 atccacaacccctgagcactcgt 2208  
 |||||  
 Db 2281 atccacaacccctgagcactcgt 2340  
 QY 2209 ggaagaacccctctcaagctggtctcaactcaagaacatttgggacaaaacccctggacagcca 2268  
 |||||  
 Db 2341 ggaagaacccctctcaagctggtctcaactcaagaacatttgggacaaaacccctggacagcca 2400  
 QY 2269 tggcagaagaagagcccttggacggcccgccagagctaaagaacagagagaagaataaactgtct 2328  
 |||||  
 Db 2401 tggcagaagaagagcccttggacggcccgccagagctaaagaacagagagaagaataaactgtct 2460  
 QY 2329 tctctactcagctgagcacaactcttctagt 2388  
 |||||  
 Db 2461 tctctactcagctgagcacaactcttctagt 2520  
 QY 2389 ccaacacatcaacactgtctctctcccaagtgatctgtatcttctagtcttcaattatct 2448  
 |||||  
 Db 2521 ccacacacatcaacactgtctctctcccaagtgatctgtatcttctagtcttcaattatct 2580

QY 2449 tcttttgatgatagacactatataaatttccatttggaaatttccaatgtgatactia 2508  
 |||||  
 Db 2581 tcttttgatgatagacactatataaatttccatttggaaatttccaatgtgatactia 2640  
 QY 2509 gtttaaatagacaggtttggaactgtgtgagactgtacttataataacttaaccgaca 2568  
 |||||  
 Db 2641 gtttaaatagacaggtttggaactgtgtgagactgtacttataataacttaaccgaca 2700  
 QY 2569 aagatcatcatctgt 2628  
 |||||  
 Db 2701 aagatcatcatctgt 2760  
 QY 2629 ttccagtgatgacaattgt 2688  
 |||||  
 Db 2761 ttccagtgatgacaattgt 2820  
 QY 2689 accttgaaatagctcccaagctgtgataatgtggcaatttggaaatttgaactttag 2748  
 |||||  
 Db 2821 accttgaaatagctcccaagctgtgataatgtggcaatttggaaatttgaactttag 2880  
 QY 2749 atgatactgt 2808  
 |||||  
 Db 2881 atgatactgt 2940  
 QY 2809 atgaacttgatctgt 2868  
 |||||  
 Db 2941 atgaacttgatctgt 3000  
 QY 2869 tcttgaggaactgt 2928  
 |||||  
 Db 3001 tcttgaggaactgt 3060  
 QY 2929 ctgcaagatgatttccagaagctgtcgaatactgttlaatcatctccatlaagggctgaa 2988  
 |||||  
 Db 3061 ctgcaagatgatttccagaagctgtcgaatactgttlaatcatctccatlaagggctgaa 3120  
 QY 2989 agaatacctacgtctctgtatatacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3048  
 |||||  
 Db 3121 agaatacctacgtctctgtatatacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3180  
 QY 3049 gaagt 3108  
 |||||  
 Db 3181 gaagt 3240  
 QY 3109 ttaagtaactgtctgt 3168  
 |||||  
 Db 3241 ttaagtaactgtctgt 3300  
 QY 3169 aaacttaagctgt 3228  
 |||||  
 Db 3301 aaacttaagctgt 3360  
 |||||  
 Db 3301 aaacttaagctgt 3360

RESULT 4  
 AF061940  
 LOCUS AF061940 3142 bp mRNA PRI 04-MAR-1999  
 DEFINITION Homo sapiens staufen protein (STAU) mRNA, alternatively spliced,  
 complete cds.  
 ACCESSION AF061940  
 VERSION AF061940.1 GI:4335950  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 3142)  
 Wickham,L., Duchaine,T., Luo,M., Nabl,I.R. and Desgroselliers,L.  
 Mammalian staufen is a double-stranded RNA and tubulin binding  
 protein which localizes to the rough endoplasmic reticulum  
 Mol. Cell. Biol. 19 (3), 2220-2230 (1999)  
 JOURNAL MEDLINE  
 99147057  
 REFERENCE 2 (bases 1 to 3142)  
 Desgroselliers,L., Wickham,L. and Luo,M.  
 AUTHORS

ORGANISM Homo sapiens  
 Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:  
 Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.  
 REFERENCE 1 (bases 1 to 3349)  
 AUTHORS Wickham, L., Duchaine, T., Luo, M., Nabl, I. R. and Desgroselliers, L.  
 TITLE Mammalian staufen is a double-stranded RNA and tubulin binding  
 protein which localizes to the rough endoplasmic reticulum  
 JOURNAL Mol. Cell. Biol. 19 (3), 2220-2230 (1999)  
 MEDLINE 99147057  
 REFERENCE 2 (bases 1 to 3349)  
 AUTHORS Desgroselliers, L., Wickham, L. and Luo, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900  
 Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada  
 FEATURES  
 Source Location/Qualifiers  
 1..3349  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="20"  
 /map="20q13.1"  
 1..3349  
 /gene="STAU"  
 /gene="STAU"  
 /feature="RNA binding protein"  
 /note="alternatively spliced transcript"  
 /codon\_start=1  
 /product="staufen protein"  
 /protein\_id="AADI7533.1"  
 /db\_xref="GI:4335953"  
 /translation="MKGKKRPMKVPDYSRMSTVYNNMGAYPPRYFPVPPPL  
 LYVELSVGGQGFNGKGTROAKHDAAKRLRIQNPRLPERLEVNRSEERNLAK  
 SEISOVEIALKRNLVNFEPARSGPRPHKNPTKYKSYSGFVEGEKSKIKKNA  
 AIAVLELKLPLPAVERVAKPRIKKTKPLVKPQTSPEYGGINPISRLAQIDQAK  
 EKEEYVTLERGLPRRREVMQVGNHNAEGTNNKKAARNAEMLEILGKVP  
 QROPKPAKLSSEKTPIKKPDGKRTVFEEGSGDENGTSNKEDEFKMPYLSHQLPA  
 GILPMVEVAOVAVSGOHHKIDFTRAAPNPAKATVAMIKRELIGSTPTAEITLK  
 NNTSSGHVPHPLRPSSEOLDYLSRVGGFOVEYKDFPKNNKNEFVSLINCSOPLLS  
 HGICKDYSECHDMAALNLIKLSLDOSTEMPTKNGPKMSVCGRC"  
 BASE COUNT 913 a 838 c 777 g 821 t  
 ORIGIN  
 Query Match 95.5%; Score 3071.8; DB 10; Length 3349;  
 Best Local Similarity 96.0%; Pred. No. 0; Mismatches 2; Indels 132; Gaps 1;  
 Matches 3215; Conservative 0; Mismatches 2; Indels 132; Gaps 1;

Db 361 CCGATAGCTGAGATTACAGAGTTATTAACTTAACCTTACACTTCAGAACTGACAAAGA 420  
 Oy caactcttctcgtgaagccctctttaaagaagaataccctactgtagaac 348  
 Db 421 CAACATTTCTCTGGAGAGCCCTCTTTTAAAGAAAGCAATACCCCTACTGTAGAAC 480  
 Oy taaatgacactggaacttgtaaaaaaaccaatgataagccttgaacccctact 408  
 Db 481 TTAATGCACTGTGATGAACCTTGAAAAAACAATGATATAGCCTGTGACCTTACT 540  
 Oy ctgcgtagtgcacactaactaacaatgagagaggtgctatccccaggtact 468  
 Db 541 CTCGATGAGTCCACATATACTACAACTAGAGAGGTGCTTATGCCCGGAGTACT 600  
 Oy ttaccctatccagttccaccttacttataagtggaacttcgttggagagacg 528  
 Db 601 TTATACCATTTCCAGTTCCACCTTACTTATCAAGTGAATCTTCTGTGGAGAGACAGC 660  
 Oy aatttaatggcaaggaagacaagacagagcttggaagaacagatgctgtgccaagcgt 588  
 Db 661 AATTTAATGGCAAGGAAGACAGACAGCTGCCAAGACAGATCTGCTGCCAAGCGT 720  
 Oy tgaagatccttgcaagatgagccctgcagagaggtctggaagtgaatggaagaaatccg 648  
 Db 721 TGAGGATCCTCGCAGATGAGCCCTGCGCAGAGAGCGCTGAGGTGAATGGAAGAAATCCG 780  
 Oy aagaagaataatcctaataatcgaataagtcaagtgttggagatgcaacttaagga 708  
 Db 781 AAGAAGAAATCTCAATTAATCTGAAATGAAGCAAGTGTGAGATTCACACTTAAACGA 840  
 Oy acttccgtgaattcgaagtgtgcccggagagatgagcccaacccaatgaagaacttg 768  
 Db 841 ACTTCTCTGTGATTTCCAGGTGCGCCCGGAGAGTGGCCACCCACATGAAGAACTTGG 900  
 Oy tgaaccaagtttcggtctggagagttgttggggagaggtgaagggaaagcaagaagattt 828  
 Db 901 TCACCAAGTTTCGGTGTGGAGGTTGTGGGGGAAGTGAAGCAAGAAAGAAATTT 960  
 Oy caaagaataaaatgcgcacatagctgttcttgaagagctgaagaagttacggccctgctg 888  
 Db 961 CAAGAAATAATGCCCCCATAGCTGTTCTTGAGAGACTGAAGATGTACCGCCCTGCTCG 1020  
 Oy cagttgaacgaatgaagccctagaatcaaaagaagaacaaacccatagtcaagccacaga 948  
 Db 1021 CAGTTGAACGAGTAATAGCTTCAATCAATAAAGAAACAAACCCATAGTCAAGCCACAGA 1080  
 Oy caagcccaagataatgcagagggatcaatccgaatagccagctgcccagatccagcag 1008  
 Db 1081 CAAGCCCAAGATATAGCCAGAGGGATCAATCCGATTAGCCGACTGACCACAGATCCAGAGG 1140  
 Oy caaaaaagagaagagagcagaagtacacgctctcaacagagagaggtctcccgcgca 1068  
 Db 1141 CAAAAAGAGAGAGAGGACAGAGTACGCTCTCACAGAGGAGGAGGCTCCGCCGCCGA 1200  
 Oy ggaagtcttgatgacagtgaaagttggaacacacacacacacacacacacacacacac 1128  
 Db 1201 GGGATTTGTGATGCAAGGTGAAGGTGAACACACACACACACACACACACACACACAC 1260  
 Oy agaagtgagccaagcgaatgacgcgaagaaatgacttggttttcaaaagttcc 1188  
 Db 1261 AGAAGGTGGCCAAAGCGCAATGACCGAGAAACATCTGTGAGATCTTGTGTTCAAAAGTCC 1320  
 Oy cgaacgagcagcccaacaaaccgcacactcaagtcaagtgagagagagacacccataaagaac 1248  
 Db 1321 CGCAGCGGAGGCCCAACCAACCCGACATCAAGTCAAGAGAGAGAACCCCAATAAAGAAAC 1380  
 Oy cagggatggaagaaagtaaccttttgaaccctggcctctggagatgaaatggagacta 1308  
 Db 1381 CAGGGATGGAAGAAAGTAACCTTTTGAACCTGTGCTGTGGATGAATATGGAGACTA 1440  
 Oy gtaataaagagatgagttcaagatgacctatcaatgaatcaagcagctgacctgtgaa 1368  
 Db 1441 GTAATTAAGAGAGATGATGATCAAGATGCTTATATAGTATCATCAGAGCTGTGCTGTGAA 1500



OY	1231	agacacccataagaacaacccaggatcggagaagaagtaacccctttttgaacctgctcgt	1230
Db	1199	AGACACCCATTAATAAACAACA - GGGAATGGAGAAAGAAATAACCTTTTAAACCTGGCTTG	1257
OY	1291	gggaatgaataatggagacagtaataaagagatgaatcagatcagatcgtctatcctaagatc	1350
Db	1258	GGGATGAATAATGGACATAGTAAATAAAGAGATGAGTTCCAGATGGCTTATCTAACTATC	1317
OY	1351	agcagctgcgctcgtggaattcttcaccaatgctgcggagatcgcggcaggctgtaagatla	1410
Db	1318	AGCAGCTGCCTCGTGGAAATCTTCCCATGGTGGCCAGGTGCCACAGGCTGAGAGATTA	1377
OY	1411	gtcaagagatcaacaccaaagatttaacaggcgagcttcogaaatccctccaaagccacg	1470
Db	1378	GTCAAGGACATCAACACCAAGATTTTACAGAGGACACTCCGAAATCCTCCCAAGGACAGG	1437
OY	1471	taactgcacatgaatgcccgagcttcttatggggagacctcggccacagccgagacca	1530
Db	1438	TAACTGCCATGATAGCCCGAGATTTGTTATG6GG6CACCTCGGCCACAGCCGAGACCA	1497
OY	1531	cttaagaataaactcctcttcagcgccagctaacccatggaactctcaacgaagacctcg	1590
Db	1498	TTTTAAAGAAATAACACTCTTTCAGGCGACAGTACCCTAGACCTGCTCAAGACCCCTCG	1557
OY	1591	agcaactcggactatcttccaagctccagggatccaggttgtaatacaagactcccca	1650
Db	1558	AGCAACTCGGACTATCTTTCCAGAGTCCAGGAGTTCAGGTTTAAATACAAACCTCCCCA	1617
OY	1651	aaaacaacaagaagaaattgtaacctataatgacctgcctcagccactcgtatca	1710
Db	1618	AAAAACACACAGACGAATTTGTATCTCTTATCATATTTGCTCTCTAGCCACCTCTGATCA	1677
OY	1711	gccatgglatcggcaagagatgtagctccctgcacatgatatgctgcgtctgacaactaa	1770
Db	1678	GCCATGGATATCGGCAAGAGTGTGAGTCCGACATGATATGAGTGGCTGGTGAACATCTTAA	1737
OY	1771	agttgctgctcgtgagttggaaccaaaagtaacagagatgcacaaacagagaacggacca	1830
Db	1738	AGTTGCTGCTGCAATTTGGACCAACAAAGTACAGAGTGCACAAAGAACAGGAACGGACCA	1797
OY	1831	tgctcgtgctgtaggaagtgctgaaccttctcggccatgaacatataaattccaca	1890
Db	1798	TGCTCTGCTGTGGAGAGTGTGAACCTTTTCTGGCCATGACATTATTAATCCACACA	1857
OY	1891	tataactgaaataactcgaacactgctcttgaaaaattggaattctcgaactccagtggg	1950
Db	1858	TATATACTGAAATAACTGAAACGTCTTGAATAATTGGAAATTTCGATACCTCCAGTGGG	1917
OY	1951	ccgagagacacggtgggtlaaaggaatgtggcgacgacgaggaagaacaacagaacaag	2010
Db	1918	CCGAGAGACACGCGTGGTAAAGATGTGGCGACGACAGGAGAGAACACAGAAACACAG	1977
OY	2011	gagcggcgtctgggc - ggcctggaactctgtctggggatctgtgtgtagtggccactcgttac	2069
Db	1978	GAGCGGCTGTGGCCGGGCTGGACGTGTGCTGGGGTTTGTGTGATGGCCACTCGGTGACC	2037
OY	2070	tgaggcttcctcaacgaaatagcagctgctcttgaggagaagagcgctgcacgacagctg	2129
Db	2038	TGGCGGTCCCTAGCCANTAGCAGCTGCTGCTGTGGGGAAGAGGGCTGCCACGACGTGCT	2097
OY	2130	tctcccgagacacagacatccacaacctgggcaacctcgttglttgltctttttctc	2189
Db	2098	TCTCCCGGAGACACAGCATGATCCACACCCCTGGGCACTCCGTGTGTTTGGCTTTTATTC	2157
OY	2190	cctgtgtgaagaagaacggcagccgacctctcaagtgtgcttaactcaaacatctg	2249
Db	2158	CTCTGTGTGAAGAAGAAACGGCACGACCCCTTCTCAAGTGGGCTCTACTCAACACATTGG	2217
OY	2250	gacaaacccctggacagcatalgacagagagagccttgacggccgacagctlaaaagca	2309
Db	2218	GACAAACCTCGACAGCATGCCACAGAGAGAGGCGTTTGAACGGGCCCCAGACTTAAAGCA	2277
OY	2310	ccagagagaataatcaaatgctctactactacgctgaccccaactttctagtgctgcaagcc	2369

Db	2278	CCAGAGAAAATGAAATGCTTCCACTACGCGTGACCCAACTTTTCTAGTGTGCACGGCC	2337
QY	2370	CCACCAACCCTCGCAgTlaccacaaccatcacacgctcttctctccttcaacagTgacTgt	2429
Db	2338	CCACCAACCCTCGCAgTlaccacaaccatcacacgctcttctctccttcaacagTgacTgt	2397
QY	2430	attcttagtTcaTtaTttcttcttgatTgataTgacataTaaattTcaattTgag	2489
Db	2398	ATTCTTAGTTCATTATTTTCTTTTGTATGATATGACACTATATATAAATTTTCTATTGTAG	2457
QY	2490	aatttctcaattTgatctgTttaaataagacagTttgTgaactTgtctgagactgactt	2549
Db	2458	AATTTCTCAATTTGTATCTTAAATAGACACATTTTGGAACTTTGCTGTGAGCTGCATTT	2517
QY	2550	atcaataactTcaacgcacacaagaTcaTcaTcatTgTgTtTgagacatttcttaTtc	2609
Db	2518	ATCATATATCTTACCGACAAAGATCATATCCATGTATGTGTTAGACATTTTATTTTC	2577
QY	2610	attgacTaaacccagagacagTtctcaTgTatTgcaaatTgtTgtccctTgtTlcaagctgaa	2669
Db	2578	ATTGACTAACCCAGCGACATTTTCAGTGAATGCAAAATTTGTGCCCTGTGGTTCCAGCTGAA	2637
QY	2670	caagtctgTactttcaaaactTgaaataagTctcccaagTctgTatTaaatTggacaatt	2729
Db	2638	CAGTCTCGACATTTCAAAAACCTTGAAATAGTCTCCACAGTGTATTAATTTGGACAAAT	2697
QY	2730	taggaaattTaaacttTaaTgatTcaTctTgtTlcaTcttTaaTlcaattTlcaTcttTgt	2789
Db	2698	TAGGAATTTTAAACTTTTACATGATCATTTTGTGTTCCATTTTATTTCATTTTACTTTTGT	2757
QY	2790	taatTgcaaacagagactTaaatTgaactTgacTctTgtTlTaaagatTatTaaaaacatt	2849
Db	2758	TATGCAAAACAGAGACTTAATATGAACCTTTGATCTCTGTTTAAAGATTATTTAAACATTT	2817
QY	2850	gtgtatctataataTatgtTcttTggagactTtagcttTcaactTcaacgcgataTgattc	2909
Db	2818	GGTATCTTATACATATATGCTCTTGGAGACTTTACCTTTACCTACATCAACGATATATCTT	2877
QY	2910	ccatTtagTcaTataaaactTgcagagTgattTtccagagTgtctgatactTgTaaTtTac	2969
Db	2878	CCATGTAGTCCATATTAACCTCGACAGTATATTTTCCAGAGTGTCTCATACTCTTAAATTAAC	2937
QY	2970	atctccatTtagggctTgaaaaagaatTgacctagTtctctTgatacagctTgtTgtctTgaa	3029
Db	2938	ATCTCCATATAGGGCTGAAAGAAATGACCTACGTTTCTGTATACAGCTGTGTTCTTTTGA	2997
QY	3030	tgTtgtTgtTactTgtTacaagaagTgtTgtcaactTgagTgtctTgtgtTgtTgcTaaTgaa	3089
Db	2998	TGTTGTGTACTGTGTACACAGAAAGTGTGTACACTGAGGCTGTGCTGTGTGCTATGGA	3057
QY	3090	aaactTgtTgacctTggaTtTaaTgactTgtTccatTcaattTgtTtcaagctTggaatttctc	3149
Db	3058	AACCTGTAGTCCCTCGGATTTAAGTACTCTCTTCCATTTATTTGTTTACCGCGGAATTTTTC	3117
QY	3150	tcaccaTggaatTgTaaTgaTaaactTaaTgtTgtTgtTcaTcaataaTatTgtTaaTactaa	3209
Db	3118	TCCCATGTGAATGTAAAGTAAAGACTTAAGTGTGTGTTCATCAATTAATGAATACTTAAAA	3177
QY	3210	aaaaaaaa 3217	
Db	3178	AAAAAAAA 3185	
RESULT 3			
AF061941	AF061941	3349 bp	04-MAR-1999
LOCUS	AF061941	3349 bp	04-MAR-1999
DEFINITION	Homo sapiens stauflen protein (STAU) mRNA, alternatively spliced,		
ACCESSION	AF061941		
VERSION	AF061941.1	GI:4335952	
KEYWORDS	human.		

RESULT	3
AF061941	
LOCUS	3349 bp mRNA PRI 04-MAR-1999
DEFINITION	Homo sapiens staufen protein (STAU) mRNA, alternatively spliced
ACCESSION	complete cds.
VERSION	AF061941
KEYWORDS	AF061941.1 GI:4335952
SOURCE	human.



[illegible][illegible]

Db	901	TTAAAGCTTAGAATAAAAAGAAAAAACCCCTAGTCAAGCCACAGACAGCCAGAT	960
Qy	961	atgyccaagggaatcaatccgattagccgacttgcgccagatccgacagycgaataaagagga	1020
Db	961	ATGGCAGGGGATCAATCCGATTAGCCAGTGGCCAGATTCAGCAGCGCAAAAAAGAGA	1020
Qy	1021	aggagccagatatacgcctcccaagagcgcctccgcgcgcaaggattgtga	1080
Db	1021	AGGAGCCAGAGTATACAGCTCTCTCCACAAGGAGGAGCCCTCCCGCGCGCAGGGATTGTGA	1080
Qy	1081	tgcagtgaaagtgtggaaaccacactgcgaagaagacgggcaccaagaaggtggcca	1140
Db	1081	TCGAGGTAAAGGTGGAAACACACTGCACAAGAAAGAGGACCAACAAGAGGTGGCCA	1140
Qy	1141	agcgcaatgcagccgagagacatgctgtgagatccctgtgttccaagtcccgcaagcgagc	1200
Db	1141	AGCGCAATGCAGCCGAGAGACATGCTGTGAGATCCCTGTGTTCAAAGTCCCGCAGCGCAGC	1200
Qy	1201	ccaccaaaccgcacatcaagtcaagagagaagacaccaataagaacccaagggatgaa	1260
Db	1201	CCACCAAACCCGCACCTCAAGTCAGAGAGAGAACCCCTAAAGAAACAGAGGAGTGGAA	1260
Qy	1261	gaaagtaacctttttggaacctgcctctgggagatgaataatgggactgtataaagg	1320
Db	1261	GAAAGTAACTTTTTTGAACCTGCGCTCTGGGATGAATAATGGGACTAATAATAAGAGG	1320
Qy	1321	atgagttcaagatgacctatcctaagatacagctgcctgctgtgaattcttcccatgg	1380
Db	1321	ATGAGTTAAGGATGACCTTATCTTAAGTATATAGACAGCTGCTGCTGTGAATTTCTCCCATGG	1380
Qy	1381	tgcaggagtcgcagcagcgtgtgaggattgataagaagaaatcacaccaaagatttacc	1440
Db	1381	TGCCGAGGTGCGCCAGCGCTGTAGAGATTGTCAAGGACATCACACCAAGATTTTACCA	1440
Qy	1441	gggcagcctccgaatccctgccaagycgaagtaactgcctatgatagccgagagttgtgt	1500
Db	1441	GGGCGACTCCGAATCTCGCCAAAGGCCAGGTAACTCCATGATAGCCCGCAGAGTGTGTG	1500
Qy	1501	atgagggaactgcgccacagccgagaacatttaagaataacatcttccaaggccag	1560
Db	1501	ATGGGGGACCTCGCCCAACAGCCAGACCATTTTAAAGATATACATCTTTACGGCCACG	1560
Qy	1561	taccocatgacactctacagagacccctctgaagaacttgaactatcttccagagatccag	1620
Db	1561	TACCCATGAGACCTCTCACGAGACCCCTGAGCACTGGACTATCTTCCAGAGTCCAGG	1620
Qy	1621	gattccaggttggatatacaagaacttcccaaaaacaagaagaattgtatctcta	1680
Db	1621	GATTCCAGGTGGAATACAAAGACTTCCCAAAAACAAGAAGACGATTTGTATCTCTTA	1680
Qy	1681	tcaattgtctctcagaacacactctgaagacatgatatcggcaagagatggagatctct	1740
Db	1681	TCAATTGTCTCTTCAGACACTCTCTGTACTGACGCATGTATCGCAAGAGATGGAGTCTCT	1740
Qy	1741	gocatgatalagctgcgcctgaacatcttaagtgtgtgtcgtagtgtgacaaccaaatga	1800
Db	1741	GCCATGATATGGCTGGCCTGAACATCTTAAGTTGGTGTCTGATTTGACCAACAAGTA	1800
Qy	1801	cagagatgccaagaacaggaacgagacaaatgtctgtgtgtgtgagagtgatcgaacctt	1860
Db	1801	CAGAGATGCCAAGAACAGGAACGAGCAATGCTGTGTGTGGAGGTGCTGAACCTTTT	1860
Qy	1861	ctggcactgaaccattataaattcccaacatatacttgaataaactctgaactctgtga	1920
Db	1861	CTGGCCATGAACCTTATAAATTCCAACATATATCTGAATAATCTGAAACTGCTTTGA	1920
Qy	1921	aaatttgaattcttgaatacccaagtggccgagagacaggtgtgttaaagaatgtgg	1980
Db	1921	AAATTTGAATTTCTGTACTCTCCAGTGGGCCAGAGACACGGTGGGTAAAGATGTGG	1980
Qy	1981	cagcagcagggaaagacacagaaacaaagagcgctgtgtgtccgcgtcgtgactgtgtg	2040

Dh	1981	CACGACGCGGGGAACAACAAGAAACACAAAGAGGCGGCTGTGGCGGCGTACGACTGTGGTG	2040
Qy	2041	gggtctgtgtgtagtgcgaactcggtagccctgycgtccctacgcgaatagcagctgcgtgt	2100
Dh	2041	GGGTTTGTGTGATGGGCGCACCTGGGTGACCTGGGGGCTCCCTACGCAATATGACACTGCTGT	2100
Qy	2101	ggggagaagaagcctgcgccagcgcagctcgtgtcttcgcgggacaccagcagatccaaacctg	2160
Dh	2101	GGGGAAAGAGGGCGCCGACACGACCTGGTCTCCCGGACACAGTACAGTCTCACACCTGG	2160
Qy	2161	ggcaacctcgggtgtgtgtctttttccacctgtggaagaagaacgcacgcacct	2220
Dh	2161	GGCACTCTCCGTGTGGTGGTCTTTTTTCCCTGTGTGAAGAAGAAACGGCACGACCTCT	2220
Qy	2221	ttccaagcttgcctcaactcaagacatctgggacaacaacctgcgaacgcaltgcgaagaag	2280
Dh	2221	TCTCAAGCTGGCTCACTGACGACATTTGGGACAACCTGTGAGAGCATGGACGATGGCAGAGAG	2280
Qy	2281	ggcctttgacggcgcccccagagctlaaagcacacagagaaaatcaatgtctctctcagcg	2340
Dh	2281	GGCTTTGACCGCGCCCGACACTAAAGACACACAGAGAAATCAAAATGCTTCTTCACTCACGCG	2340
Qy	2341	tgaaccaacttttctagtgtgcacagcgccccaccacctctgcagtaaccaacaataac	2400
Dh	2341	TGACCAACTTTTCTAGTGTGGCCACGGCCCCACACCTCTCTGAGTACCCACACCATCAC	2400
Qy	2401	cactgccttctctctccaacagtgatcgcgtatctcttagttctcaatcttctcttgatga	2460
Dh	2401	CACGTCTTTCCTTCCCAAGGTATGTGATTTCTTGAATTTTCTTTGATTTGA	2460
Qy	2461	taigacactatataaaatcttcatcttggagaattctctcaattgtaacctgtaataaagcac	2520
Dh	2461	TATGACACTATATATAAATTTTTCATTTTGACAAATTTTTCATTTATCTAGTTAAATAGCAC	2520
Qy	2521	agtttggaaactgtctgcgaagcagctttaaataatctaaacgaagaagtaataacc	2580
Dh	2521	AGTTTGGAAACTTGTCTGAGACTGACTTTTATCATATATCTAATACCGCAAGATCATATATCC	2580
Qy	2581	alggtatgtgtgttagacatttctatctcaatbaactcaaccacgagagcttctaaagtatgc	2640
Dh	2581	ATGTGTATGTGTGTAGACATTTTATTTTCTTCATTGACTAACCCAGAGACGTTTCAGTGATGC	2640
Qy	2641	aaathtgtgtgcctctcgtgttcgaagctgaacagctcctgcgaacttcaaaaaaccttgaataag	2700
Dh	2641	AAATGTGTGCGCCTCGTGTGACGTGAACAGTCCGTGACTTTCAAAAACTTGAATAG	2700
Qy	2701	ttccccaagtgatgataaatgtggaacaattgaagaatttaaaacttagatgataatttgg	2760
Dh	2701	TCTCCACAGTGTGATATAATTTGACAAATTTAGGAATTTTAAACTTGTAGTGTACTTTGGG	2760
Qy	2761	ttccaatttcttcatcttctatcttgttaagcaaacagagacttaaaagacttcat	2820
Dh	2761	TTCCATTTTATTTTCATTTTATTTTGTTAATGCAAAACAGGACTTAAATGAACCTTTGAT	2820
Qy	2821	ctctgtttaaagaatlaataaaaaactgtgtatctatatacatalgycbctbtaggact	2880
Dh	2821	CTCTGTTTAAAGATATTTAAAAAACCTGTGTATCTATACATATGCGCTTGAAGACTT	2880
Qy	2881	agccttcaactacacacagatatgatctccatgtatgccatatataaactgcgaagtgat	2940
Dh	2881	AGCTTTCACCTACACTACAGGATATGATCTCCATGTGTGTCATATTAACCTGCAGAGTAT	2940
Qy	2941	tttccaagatgcctgcatacagtttaattacactccaacttaggcctgaagaagaatagcaccac	3000
Dh	2941	TTTCCAGAGTCTCGATACGTGTTAATTTACATCTCCATTAGGGCTGAANAAGATGACTTAC	3000
Qy	3001	gttctcgtatatacagctgtgtgtcttcttgatgtgtgtcttacaagaagtggtgtga	3060
Dh	3001	GTTTCTGTATACACTGTGTGTGTGCTTTTGATGTGTGTACTGTACACAGAAAGTGTGTCA	3060
Qy	3061	ctgagagctctgctgtgtgtccgtataggaaaacctggttagccctgcgagttaagtactgct	3120
Dh	3061	CTGAGAGCTCTCGTGTGTGTCCGTATAGGAAAACCTGTGTACCTCTGGAGAGTTAAGTACTGCT	3120

RESULT	1	ALIGNMENTS
AF061938		
LOCUS	AF061938 3217 bp mRNA	PRI 04-MAR-1999
DEFINITION	Homo sapiens staufen protein (STAU) mRNA, alternatively spliced, complete cds.	
ACCESSION	AF061938	
VERSION	AF061938.1	GI:4335946
KEYWORDS		
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	1 (bases 1 to 3217)	
TITLE	Wickham,L., Duchaine,T., Luo,M., Nab,I.R. and Desgroseillers,L. Mammalian staufen is a double-stranded RNA and tubulin binding protein which localizes to the rough endoplasmic reticulum Mol. Cell. Biol. 19 (3), 2220-2230 (1999)	
JOURNAL	99147057	
MEDLINE	2 (bases 1 to 3217)	
REFERENCE	Desgroseillers,L., Wickham,L. and Luo,M.	
AUTHORS	Direct Submission	
TITLE	Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900	
JOURNAL	Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada	
FEATURES	Location/Qualifiers	
source	1..3217	
gene	/organism="Homo sapiens"	
CDS	/db_xref="taxon:9606"	
	/chromosome="20"	
	/map="20q13.1"	
	1..3217	
	/gene="STAU"	
	363..1853	
	/gene="STAU"	
	/function="RNA binding protein"	
	/note="alternatively spliced transcript"	
	/codon_start=1	
	/product="staufen protein"	
	/protein_id="AA017530.1"	
	/db_xref="GI:4335947"	
	/translation="MKLGKPMYKRPDPYSRMOGNYNNMRGAVPRPYFPVPVPPV	
	LYGVLSVGGQGFNGKGTROAKHDAAKALRILONEPLRPERLVNGRSESEENLNN	
	SEISQVEIETALKRNLVPNFEVARSGRPHMKNPPTKYSVGEFVGESEKSKISKKN	
	AIATVLEELKRLPPLPAVERVAPRIKTKTKPLVKTQTSYEGQINPISRLAIOIQAKK	
	EKEPEYVILTEGRPLRRRERPMQVYQVNHTEEGTGNKVKAKRNAENMELTGKRV	
	AC0055879 Homo sapi	
	AK001576 Homo sapi	
	ALJ390198 Homo sapi	
	AC068845 Homo sapi	
	AK023314 Homo sapi	
	XB3973 H.sapiens m	
	ME9111 D.melanog	
	AC004336 Drosophi	
	AC017563 Drosophi	
	AE003800 Drosophi	
	AF225924 Drosophi	
	AF225925 Musca dom	
	AC027018 Homo sapi	
	166494 Sequence 14	
	166494 Sequence 14	
	AC023436 Homo sapi	
	AC078996 Mus muscu	
	AC010758 Homo sapi	
	U57949 Genomhabd	
	AC023197 Mus muscu	
	AL109755 Human DNA	
	AC011801 Homo sapi	
	AC015716 Homo sapi	
	AL137078 Human DNA	

Query Match	Best Local Similarity	99.9%	Score 3213.8;	DB 10;	Length 3217;
Matches 3215;	Conservative	0;	Mismatches	2;	Indels
					Gaps
					0;
QY 1	actcttcgcgggctcgcgagcgctcgagcgcttcacagcgtttgcgggcgctgagcgt	60			
Db 1	ACTTCTTCGTGGCGGGCGCGCGCGCTGAGGGCTTCTTCAAGGTTTGGCGGGCGCTGGCGCT	60			
QY 61	ctctctcgctccgacttcctcttgacgcgctcccccccccgcccgcgcgcgccgcctc	120			
Db 61	CTCTCTCGGCTCCGCTTCTTGTACCGGCTCCCCCGCGCGCGCGCGCGCGCTC	120			
QY 121	ctccacgycgacatccgctctctccctcccttcgctctctctctctctctctctct	180			
Db 121	CTCCACGGGCACTCGGCTCTTCCCTCCCTTGCTCCCTTCTTCTTCTTCTTCTTCTTCTT	180			
QY 181	ctctctccctccctccgcgccgacgccccgagacgccccgcgcgcgcgctcgagc	240			
Db 181	CTTCTCTTCCCTTCCCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	240			
QY 241	agcagcgagagttttaaaccatcaccctccagagctgaactgaacaaagcaaatgtcc	300			
Db 241	AGCAGCGAGTTTATTAAACCTTAACTTCCAGACTGGAACAAAGACATTTGTTCC	300			
QY 301	tggaacgcccccttlttaaaaaagaagcataccctactgtagaactaaatgcactgt	360			
Db 301	TGGAACGCCCTCTTTTAAAAAGAAAGCATTAACCCCTACTAAGAACTAAATGCACTGT	360			
QY 361	gcatgaacttgtaaaaaaaccaatgataaacctgttgacccctactctcgatgcagt	420			
Db 361	GCATGAACCTTGGAAAAAACCAATGTATTAACCTGTGTACCCCTTAACCTTCGGATGCACT	420			
QY 421	ccacctataacataacatgagagagtgctatacccccggagctacttaccatctc	480			
Db 421	CCACCTATTACTACAACTAGAGAGAGGTGCTTATCCCGGAGGTACTTTTACCATTTCC	480			
QY 481	cagttccaccttactttaaactgaaacttctgttggaagaaagcaatttaaattgca	540			
Db 481	CAGTTCCACCTTACTTATTAAGTGAAGAACTTCTGTGGGAGACAGCAATTAATGGCA	540			
QY 541	aaggaagaagaacagacagctgcgaaacacgatgctgtctgcaaaagcgttgaagctcgc	600			
Db 541	AAGGAAGACAAAGACAGAGCTGCGAAACACGATGCTGCTCCAAACGTTTGAAGATCTTGC	600			
QY 601	agaatgagccccctgcagagagcgtcgaggtgtaatgggaagagaatccgaagaagaatc	660			
Db 601	AGAAATGAGCCCTTGCACAGAGAGGCTGAGAGTAAATGGAAGAAATTCGAAAGAAATATC	660			
QY 661	tcaatcaatctgaataagtcaagtggttgagattgcaacttgaagaagaactgctgtga	720			
Db 661	TCAATTAATCTTAATTAAGTCAAGTGTGTAAGATTTGCACTTAAAGGAAGTCTGCTGTGA	720			
QY 721	atttcgaggtgccccgagagtgagcccaacccacatgaagaacttgtgacaaagttt	780			
Db 721	ATTTCGAGGTGGCCCCGGGAGAGTGGGCCACCCACATGAAGAACTTGTGACCAAGGTTT	780			
QY 781	cgggtcgaggagcttctgtggggaagtgaaagggaaagcaagaagatttcaagaagaatg	840			
Db 781	CGGTTGGGAGATTTTGTGGGGAAGGTGAAGGAAAAAGCAAGAAAGATTTTCAAGAAAAATG	840			
QY 841	cggcacaagcttcttgaagagctgaaagaattaccggcccttgctcgaattgaagag	900			
Db 841	CGGCATACCTTCTTTGAGAGCTGAAAGACTTACCGCCCTGCTGACAGTTGAACGAG	900			
QY 901	taaagccttgaatcaaaaaaagaaacccatagtcgaagccacagaacaagcccaagat	960			

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd

OM nucleic - nucleic search, using sw model

```
Run on:      April  5, 2001, 08:30:54 ; Search time 24649.1 seconds
              (without alignments)
              667.928 Million cell updates/sec
```

Title: US-09-316-048-3  
 Report Number: 2017

Sequence:    1 acttcctgcgcygctgcgg ..... taataactaaaaaaaaaa 3217

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.00

Searched: 1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Maximum DB seq length: 20000000000

Post-processing:	Minimum Match 08
	Maximum Match 100%

Listing first 45 summaries

Database :

1:	gEmbEmb1.*
2:	gb_ba1.*
3:	gb_ba2.*
4:	gb_om.*
5:	gb_ph.*
6:	gb_p11.*
7:	gb_p12.*
8:	gb_p11.*
9:	gb_p12.*
10:	gb_p12.*
11:	em_fun.*
12:	em_hum1.*
13:	em_hum2.*
14:	em_in.*
15:	em_in.*
16:	em_ov.*
17:	em_ov.*
18:	em_pat.*
19:	em_ph.*
20:	em_ph.*
21:	em_ro.*
22:	em_ro.*
23:	em_sy.*
24:	em_sy.*
25:	em_v1.*
26:	gb_hg1.*
27:	gb_hg2.*
28:	gb_in1.*
29:	gb_in2.*
30:	em_ba1.*
31:	em_ba1.*
32:	em_hum3.*
33:	em_hum4.*
34:	gb_p14.*
35:	gb_hg3.*
36:	gb_hg4.*
37:	gb_hg5.*
38:	gb_hg6.*
39:	gb_hg7.*
40:	em_hg2.*
41:	em_hg2.*
42:	em_hg3.*
43:	em_hum5.*

SUMMARIES					
Result	Query	Match	Length	DB	ID
No.	Score				
1	3213.8	99.9	3217	10	AF061938
2	3105.2	96.5	3194	34	AK025519
3	3071.8	95.5	3349	10	AF061941
4	3053.8	94.9	3142	10	AF061940
5	2914.8	90.6	3506	10	AF061939
6	2759.4	87.0	3066	68	HS0132258
7	1357.4	42.2	113836	34	AL133174
8	1281.8	39.8	3041	74	AF290989
9	1273	39.6	2860	74	AF061942
10	1253.8	39.0	3059	74	AF0227200
11	705	21.9	157803	35	AC011492
12	553.2	17.2	194407	54	AC068845
13	383.2	11.9	192587	70	AL158850
14	380.2	11.8	385	75	G30939
15	287	8.9	360	74	RN0010200
16	283.6	8.8	225016	52	AC063967
17	277	8.6	2968	34	AK002152
18	277	8.6	4058	68	HS09062
19	220.2	6.8	263	75	G22107
20	215	6.7	160979	52	AC055879
21	193.6	6.0	163253	68	HS08140E4



AUTHORS	TITLE	JOURNAL	LOCATION/Qualifiers
Kiebler, M.A., Henrat, I., Verbrade, P., Kohmann, M., Rotes, P., Marlon, R.M., Otin, J., and Dotli, C.G.	The mammalian staufen protein localizes to the somatodendritic domain of cultured hippocampal neurons: implications for its involvement in mRNA transport	J. Neurosci.	19 (11), 288-297 (1999)
			99088098

```

Db 60502 ACATTCTTTCTACTGACATACCAAGACCAATTTCAGTGTGCAATTCATGTACCTCT 60443
QY 2582 ggctcagctgaacaagcctcggactt-tcaaaacctgaataagtcocacagctgta 2640
Db 60442 GGTGTACTGTGAAAAAGCCCTGGACTTCTCAGAAAACTTGAAATTTTCCATCATGTGA 60383
QY 2641 taattgacacatttagaatttaacttagatgacatgcttgctccatttatttc 2700
Db 60382 CAAATGTGGCCATTGTAGCAATTCACAACTTCA---ATCATTTGGTCTCTTTATATTTT 60327
QY 2701 atttattttgttaatgca--aacagagacttaatgaacttgatctcgttttlaag 2758
Db 60326 ATTTTATTTTGTGTGTATGGGACCTAACAATAAATTTGATCTTGTTTTAAAG 60267
QY 2759 attactaaacaacttgatctatcatatgagcctcttgagagacttgaacttaca 2818
Db 60266 ATTTTAAAAAATTACGCTGTATGACATATATGCTTTCAGACCCACAGCTTTCATATG 60207
QY 2819 ctacagatatgactcctcatgtagtccatataaactgacagagatgttcc-agagtc 2877
Db 60206 CTGTAGGATATGTGCTCTGTGTATGACATATGACATTCGACGTTAATTTCTTGATGTC 60147
QY 2878 tcgatactgtaattacac 2897
Db 60146 TTTATACCTCTATTTTCTC 60127

```

```

RESULT 14
LOCUS G30939 385 bp DNA STS 29-SEP-1998
DEFINITION DRESS5 Human DRES Homo sapiens STS genomic clone 22368 3', sequence
tagged site.
ACCESSION G30939
VERSION G30939.1 GI:1724032
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 385)
AUTHORS Banfi,S., Borsani,G., Rossi,E., Bernard,L., Guifanti,A.,
Rubboli,F., Marchitelli,A., Giglio,S., Coluccia,E., Zolli,M.,
Zuffardi,O. and Ballabio,A.
Identification and mapping human cDNAs homologous to Drosophila
mutant genes through EST database searching
Nat. Genet. 13 (2), 167-174 (1996)
96225443

```

TITLE JOURNAL MEDLINE COMMENT

Contact: Ballabio Andrea  
Telethon Institute of Genetics and Medicine, TIGEM  
Telethon Institute of Genetics and Medicine, TIGEM  
via Olgettina 58, 20132 Milano, Italy  
Tel: +39-2-215601  
Fax: +39-2-21560220  
Email: dres@tigem.it

Primer A: CTAAACATGATGCAAGCAG  
Primer B: GGCTGAAAAAGATGACTTAC  
STS size: 157  
PCR Profile:

Presoak: 94 degrees C for 2 minutes  
Denaturation: 94 degrees C for 45 seconds  
Annealing: 57 degrees C for 45 seconds  
Polymerization: 72 degrees C for 45 seconds  
PCR Cycles: 35  
Thermal Cycler: MJR PTC-200

Protocol:

Template: 25 ng  
Primer: each 1 uM  
dNTPs: each 220 uM  
Tag Polymerase: 0.025 units/uM  
Total Vol: 15 uL

Buffer: MgCl2: 1.5 mM  
KCl: 50 mM  
Tris-HCl: 10 mM  
pH: 8.3

The 5' end of the cDNA clone 22368 is homologous to the staufen  
Drosophila gene product. EST from which STS is derived: 789105.  
Forward primer is 70-1233; Reverse primer is 70-1234.

FEATURES  
source  
Location/Qualifiers  
1..385

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="20, 12.10 CR from GCT10P11 (lod >3.0)"  
/clone="22368"  
/clone\_1id="Human DRES"  
STS  
primer\_bind  
29..185  
primer\_bind  
29..49  
BASE COUNT 135 a 79 c 71 g 100 t  
ORIGIN

Query Match 12.1%; Score 380.2; DB 75; Length 385;  
Best Local Similarity 99.2%; Pred. No. 8.7e-78;  
Matches 382; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 2705 ttatttgttaatgcaaacagagacttaaalgaacttgatctctgttttaagattatc 2764
Db 385 TTTATTTTGTAAATGCAAAACAGGACTTAATGAAGACTTGTATCTGTGTAAAGATTATTT 326
QY 2765 aaaaacatgctgcatatcatatgctcttgagagacttgaacttccactacactacag 2824
Db 325 AAAAACAATGTGTATCATATCATATGAGCTCTTGAAGACTTACCTTACATACACTACAG 266
QY 2825 gatatgactcactgtagtccatataaaccctgcagagtgatcttccagagctgcaglac 2884
Db 265 GATATGATCTCCATGATGATGATCATATTAACCTGCAGAGATTTCCAGAGTGTGCATATC 206
QY 2885 tgttaatacatctccatcattagagctgaaagaatgactaagcttctgtataacgctgtg 2944
Db 205 TGTATATTACATCTCCATATAGGCTGAAAGAAATGACCTAGCTTCTGTATATACGCTGTG 146
QY 2945 ttgctcttgatgctgtgtgctacgtacacagaagtgctgcagctgagctgcgtgtgtg 3004
Db 145 TTGCTTTTGTATGTTGTGTACTGTACACAGAGTGTGTGCACTGAGGCTGTGCTGTGCT 86
QY 3005 ccglatggaacactgtagcgcctgcagagtgtaagtaactgcttccatcattgattacgct 3064
Db 85 CCGTATGGAAGCCGTGATGAGCCCTGCGAGGTAAAGTACTGTTCCATTCATTTAGTACCT 26
QY 3065 ggaattttctcccatggaatgta 3089
Db 25 GGAATTTTCTCTCCCATGGAATGTA 1

```

```

RESULT 15
LOCUS RNO010200 360 bp mRNA ROD 22-JAN-1999
DEFINITION Rattus norvegicus mRNA for Staufen protein, partial.
ACCESSION AJ010200
VERSION AJ010200.1 GI:4138427
KEYWORDS staufen gene.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 360)
AUTHORS Kiebler M.A.
TITLE Direct Submission
JOURNAL Submitted (16-AUG-1998) Kiebler M.A., Cell Biology, EMBL,
Meyerhofstr. 1, D-69117 Heidelberg, GERMANY
REFERENCE 2 (bases 1 to 360)

```

RESULT 13  
 AL158850/C  
 LOCUS  
 DEFINITION Homo sapiens chromosome 6 clone RP3-522B19, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*, 11 unordered pieces.  
 ACCESSION AL158850  
 VERSION AL158850.4 GI:9800143  
 KEYWORDS HTG: HTGS\_PHASE1, HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 192597)  
 AUTHORS Burton, J.  
 JOURNAL Direct Submission  
 Submitted (05-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clonerequest@sanger.ac.uk  
 On Aug 13, 2000 this sequence version replaced gi:9212635.  
 ----- Genome Center  
 Center: Sanger Centre  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquery@sanger.ac.uk  
 ----- Project Information  
 Center project name: dj522B19  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Sequencing vector: M13; M77815; 12% of reads  
 Sequencing vector: plasmid; L08752; 87% of reads  
 Chemistry: Dye-terminator Big Dye; 77% of reads  
 Chemistry: Dye-terminator-amersham; 22% of reads  
 Chemistry: Dye-primer Big Dye; 0% of reads  
 Consensus quality: 185625 bases at least Q40  
 Consensus quality: 188533 bases at least Q30  
 Consensus quality: 190005 bases at least Q20  
 Insert size: 191597; sum-of-ctrls  
 Insert size: 144116; 9.6% error; agarose-fp  
 Quality coverage: 5.71x in Q20 bases; sum-of-ctrls Quality  
 coverage: 8.21x in Q20 bases; agarose-fp  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 11 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 1 13837: contig of 13837 bp in length  
 \* 13838 13937: gap of 100 bp  
 \* 13938 17476: contig of 3539 bp in length  
 \* 17477 17576: gap of 100 bp  
 \* 17577 20005: contig of 2429 bp in length  
 \* 20006 20105: gap of 100 bp  
 \* 20106 32070: contig of 11965 bp in length  
 \* 32071 32170: gap of 100 bp  
 \* 32171 39772: contig of 7602 bp in length  
 \* 39773 39872: gap of 100 bp  
 \* 39873 43471: contig of 3599 bp in length  
 \* 43472 43571: gap of 100 bp  
 \* 43572 51795: contig of 8224 bp in length  
 \* 51796 51895: gap of 100 bp  
 \* 51896 178838: contig of 126943 bp in length  
 \* 178839 178938: gap of 100 bp  
 \* 178939 181919: contig of 2981 bp in length  
 \* 181920 182019: gap of 100 bp  
 \* 182020 184231: contig of 2212 bp in length  
 \* 184232 184331: gap of 100 bp  
 \* 184332 192597: contig of 8266 bp in length.  
 \*  
 Location/Qualifiers

source 1. 192597  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /clone="RP3-522B19"  
 /clone\_lib="RP3-3"  
 1. 13837  
 /note="assembly\_fragment:02383  
 fragment\_chain:1"  
 13938. 17476  
 /note="assembly\_fragment:01080  
 fragment\_chain:1"  
 17577. 20005  
 /note="assembly\_fragment:02434  
 fragment\_chain:1"  
 20106. 32070  
 /note="assembly\_fragment:03060  
 fragment\_chain:2"  
 32171. 39772  
 /note="assembly\_fragment:00945  
 fragment\_chain:2"  
 39873. 43471  
 /note="assembly\_fragment:00013  
 fragment\_chain:2"  
 43572. 51795  
 /note="assembly\_fragment:02225  
 fragment\_chain:3"  
 51896. 178838  
 /note="assembly\_fragment:01767  
 fragment\_chain:3"  
 178939. 181919  
 /note="assembly\_fragment:01637"  
 182020. 184231  
 /note="assembly\_fragment:02408"  
 184332. 192597  
 /note="assembly\_fragment:02960"  
 /note="assembly\_fragment:02960"  
 BASE COUNT 56324 a 39515 c 38313 g 55431 t 1014 others  
 ORIGIN  
 Query Match 12.2%; Score 383.2; DB 70; Length 192597;  
 Best Local Similarity 73.2%; Pred. No. 3.3e-78;  
 Matches 586; Conservative 0; Mismatches 203; Indels 11; Gaps 7;  
 QY 2102 gttcttttttccccgtgttgaaagaagaacgacccctctcaagctgctcac 2161  
 || ||||| || ||||| ||||| || |||||  
 DB 60919 GTAATTTTATTTCGCTGTGTGAAAAAAGAAATGAGACAGAAAGATTACCTGCTCAC 60860  
 QY 2162 tcagacacatgagcaaacacctgacagccatgacagagagagccttgacggccc 2221  
 || ||||| || ||||| || ||||| || ||||| || |||||  
 DB 60859 ACTGCTGACTCAAC -ACCTTGCTTTAAACATGCTGAGAAAAAGCTTAGACTGCGCTC 60801  
 QY 2222 agagctaaagaacacagagaaatcaatgcttcctcactcagcgtagcccaatttctca 2281  
 || ||||| || ||||| || ||||| || ||||| || ||||| || |||||  
 DB 60800 AAAGCTGAAACACCCGGGAAAAATTAACTGCTTCACTGTTACAAATTTACCTTCCTCA 60741  
 QY 2282 ggtgtccacggccccccctctcagtaaccacacatcaacacatccttctctcc 2241  
 || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||  
 DB 60740 GATGCTACGG -CCAGTCACTTACTGACAGTACCCACCACTAACCACTGCTTCTTCT 60682  
 QY 2342 aacagtgatctglatctctagttcattctcttcttgatgatatagaactataaa 2401  
 || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||  
 DB 60681 AATAGGATACATATTTTATTTATTTA -TTTCTTTGATTAAGTAATGATGCAATATGAA 60623  
 QY 2402 atttcattggaattctcattglatctagtttaataatagacagttggaaactgtc 2461  
 || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||  
 DB 60622 ATTTCATTGGAAGTTTCTCAGTTATCTATCTAGTATGATAGTACAGTACGAAGCTGTC 60563  
 QY 2462 tgaagctgactatataatctaacgcaagaagatcatatccatgtagttagttag 2521  
 || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||  
 DB 60562 CCAGACTGACTTATTCAGGAATCTAATGCAATGATATCCAGTATAGTGGCTGT 60503  
 QY 2522 acattttattcatgtaatacccgagcagtttcagtgatgacaattgtgtgcctct 2581





```

Db 149889 TGCTGGGATTACTGCGATGAGCCACCGTGCCGACGACATATATAAATTTTCATTGGA 149830
Qy 2414 gaattctcaatgtagtctgaatgaatagacagatttggaaacttctctagaactt 2473
Db 149829 GAGTTTCTCAGCTGATCTATATATATATGACAGGTTTGGAACTTTTCTGAGCTGACTT 149770
Qy 2474 tatcaataactcaacacgacaaagatcatatcctatgctgtagttagagacttttatt 2533
Db 149769 TATCAATTAATCTTAACCGACAAAGATCATATCCAGCTGTATGTATGAGACATTTTATT 149710
Qy 2534 catgtgtaaccacagagacagtttccagtatgcaaatgtgtccctctggttcaagttaa 2593
Db 149709 CATTTGACTAATCCGACGACATTTTCAGTGAAGCAAAATTAAGTCCCTATGAGTTACGCTGAA 149650
Qy 2594 acagtcctggaacttcaaaaaaacttgaataagctccacagctgtatgaatggagcaat 2653
Db 149649 ACAGTCTCTGACTTTTCAAAAACCTTGATATAGTCTCCACAGTTGTATTAATTTGGACAA 149590
Qy 2654 ttagaattttaaactttagatgcatcatttgcattttatatttcaattttattttt 2713
Db 149589 TTAGGAATTTTAAAGTTTAACTTATGATGATCATTTGCTTCATTTTATTTATTTTATTTTG 149530
Qy 2714 ttaatgcaaacagagactt-----aaatgaacttgcactctgtttaaagattttaa 2769
Db 149529 TTAAATGCAAAACGAGACTTAAACAAATGAACCTTGATCTCCATTTTAAATATCTTTAAGA 149470
Qy 2770 acattgtatctatcatatagatgctcttgagagacttgaacttcaactacactacagatat 2829
Db 149469 AATTCATGCTGCTATATACATACGCTCTTGAGGACTTACCTTACATACACTACAGCATAT 149410
Qy 2830 gatctccatgtagtcatatataaactgcagagagatttccacagatgctcgaactgtagt 2889
Db 149409 GATCTCCACATAGTTCATACATAAACGACAGCTGATTTTCCAGAGCTCGATACATATTA 149350
Qy 2890 attaatctccattagagctgcgaagaatgacttaactgttctgtatatacagctgtgtgt 2949
Db 149349 ATTACATCTCGATGAGGCTGGAAGATGACCTATGTTCTTTATACAGTGTGTGTGCT 149290
Qy 2950 tttagctgtgtgttactgtagtgcagagatgctgtgcactgagagc 2992
Db 149289 TTTCGATGTTGTGTTACTATACACAGAGTGTGTCACCTGGGCC 149247

RESULT 12
AC068845/c DNA HTG 04-SEP-2000
LOCUS Homo sapiens chromosome 19 clone RP11-73711 map 19, WORKING DRAFT
DEFINITION
SEQUENCE, 27 unordered pieces.
ACCESSION AC068845
VERSION AC068845.3 GI:9966312
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS Birren, B., Linton, L., Nussbaum, C., and Lander, E.
TITLE Homo sapiens chromosome 19, clone RP11-73711
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 194407)

```

# TITLE JOURNAL COMMENT

```

Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Olliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange, R., Thoman, N., Stojanovic, N., Subramanian, A., Talmas, J.,
Testa, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wymann, D., Ye, W. J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Submitted (10-MAY-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 4, 2000 this sequence version replaced g1:8247860.
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RW/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7549
Center clone name: 737_L1
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 181875 bases at least Q40
Consensus quality: 187842 bases at least Q30
Consensus quality: 190220 bases at least Q20
Insert size: 190000; agarose-fp
Insert size: 191807; sum-of-ctrls
Quality coverage: 4.7 in Q20 bases; agarose-fp
Quality coverage: 4.7 in Q20 bases; sum-of-ctrls
-----
NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1755: contig of 1755 bp in length
1756 1855: gap of 100 bp
1856 3429: contig of 1574 bp in length
3430 3529: gap of 100 bp
3530 5346: contig of 1817 bp in length
5347 5446: gap of 100 bp
5447 7338: contig of 1892 bp in length
7339 7438: gap of 100 bp
7439 10297: contig of 2859 bp in length
10298 10397: gap of 100 bp
10398 13590: contig of 3193 bp in length
13591 13690: gap of 100 bp
13691 16293: contig of 2603 bp in length
16294 16393: gap of 100 bp
16394 19065: contig of 2672 bp in length
19066 19165: gap of 100 bp
19166 22713: contig of 3548 bp in length
22714 22813: gap of 100 bp
22814 27165: contig of 4352 bp in length
27166 27265: gap of 100 bp
27266 32377: contig of 5112 bp in length
32378 32477: gap of 100 bp
32478 36572: contig of 4095 bp in length
36573 36672: gap of 100 bp
36673 41657: contig of 4955 bp in length
41658 41757: gap of 100 bp
41758 47024: contig of 5267 bp in length
47025 47124: gap of 100 bp
47125 52938: contig of 5814 bp in length
52939 53038: gap of 100 bp

```

AC011492/c  
 LOCUS AC011492 157803 bp DNA HTG 04-MAY-2000  
 DEFINITION Homo sapiens chromosome 19 clone CTB-187L3, WORKING DRAFT SEQUENCE,  
 8 unordered pieces.  
 AC011492  
 VERSION AC011492.5 GI:7690131  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 (bases 1 to 157803)  
 DOE Joint Genome Institute.  
 TITLE Sequencing of Human Chromosome 19  
 JOURNAL Unpublished  
 2 (bases 1 to 157803)  
 DOE Joint Genome Institute.  
 TITLE Direct Submission  
 AUTHORS Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 JOURNAL On May 4, 2000 this sequence version replaced gi:7458736.  
 COMMENT  
 -----Genome Center  
 Center: Joint Genome Institute  
 Center Code: JGI  
 Web site: http://www.jgi.doe.gov  
 -----  
 Project Information  
 Center Project Name: 136279, BC101302  
 Center clone name: CIT978SKB\_187L3  
 -----  
 Summary Statistics  
 Consensus quality: 152548 bases at least Q40  
 Consensus quality: 153824 bases at least Q30  
 Consensus quality: 154424 bases at least Q20  
 Estimated insert size: 156750; agarose-fp estimation  
 Estimated insert size: 157103; sum-of-contrigs estimation  
 Quality coverage: 9.52 in Q20 bases; agarose-fp estimation  
 Quality coverage: 9.5 in Q20 bases; sum-of-contrigs estimation.  
 NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 8 contrigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contrigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 1  
 2374: contig of 2374 bp in length  
 2375 2474: gap of unknown length  
 2475 4521: contig of 2047 bp in length  
 4522 4621: gap of unknown length  
 4622 7566: contig of 2945 bp in length  
 7567 7666: gap of unknown length  
 7667 12083: contig of 4417 bp in length  
 12084 12183: gap of unknown length  
 12184 18288: contig of 6105 bp in length  
 18289 18388: gap of unknown length  
 18389 63650: contig of 45262 bp in length  
 63651 63750: gap of unknown length  
 63751 111373: contig of 47623 bp in length  
 111374 111474: gap of unknown length  
 111474 157803: contig of 46330 bp in length.  
 location/qualifiers  
 1.157803  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="19"  
 /clone="CTB-187L3"  
 /clone\_lib="Caltech human BAC library B"  
 BASE COUNT 40983 a 38390 c 39640 g 38079 t 711 others  
 ORIGIN

Query Match

22.4%; Score 705; DB 35; Length 157803;

Best Local Similarity 72.5%; Pred. No. 5.3e-153;  
 Matches 1206; Conservative 0; Mismatches 90; Indels 367; Gaps 8;  
 QY 1677 gctgcgtgaacatcttaaggttgcctgtgagttggaaccaacaagaatgaagatgcc 1736  
 |||||  
 Db 150889 GCTGCCTGAACACCTTAAGTTGCTCTGAGTTGACCAACAAATACAGATGCCA 150830  
 QY 1737 agaacaggaacagaccatgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1796  
 |||||  
 Db 150829 AGAACAGGAATGATATATGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 150770  
 QY 1797 ccattataaaatcccaacatactactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1856  
 |||||  
 Db 150769 CCATTATATAATCCCAACATATACAGATATATATATATATATATATATATATATAT 150710  
 QY 1857 ttctgatactccagtggt 1916  
 |||||  
 Db 150709 TTCTGATACCTCTGT 150652  
 QY 1917 aagaacaacagaaacacag 1975  
 |||||  
 Db 150651 AACACAAACG-----GGAGGTGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 150599  
 QY 1976 tgatgccaactcgtgtgactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2035  
 |||||  
 Db 150598 TGACGGCCACTCGGTGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 150539  
 QY 2036 ggtgcgccaacagctgt 2095  
 |||||  
 Db 150538 GGTGTGCCAGCAGCTGT 150479  
 QY 2096 tgttggctcttttttccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2153  
 |||||  
 Db 150478 TGTTTGTCTTTTTCCTGT 150419  
 QY 2154 tggctcactcagacacatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2213  
 |||||  
 Db 150418 TGCTTTAG-----TTGGGACAAACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 150368  
 QY 2214 ccggccccaagagcctaaagacacagagagagagagagagagagagagagagagagag 2273  
 |||||  
 Db 150367 CCACCCCAAGAGCTTAAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 150308  
 QY 2274 cttctcagtggt 2333  
 |||||  
 Db 150307 C--TTTAGTGTGCCAGGCCCCACCACTTCTGTGTGTGTGTGTGTGTGTGTGTGTGT 150250  
 QY 2334 tctctccaacagtgatctgtatcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2388  
 |||||  
 Db 150249 TCTCTTCCAAAGAGATCTGTATTTAGTTTATTTCTTTTATTTGTATGTATGACAC 150190  
 QY 2389 ----- 2388  
 Db 150189 TATATAAATTTTCT 150130  
 QY 2389 ----- 2388  
 Db 150129 TTCTTTCTTCAAGACAAAGTCTACTCTGTACACAGGCTGAGTGACAGGACAGTC 150070  
 QY 2389 ----- 2388  
 Db 150069 TTGGCTCCTGCAACTTCTGCTCCAGAGTTCAAGCAATTTGTGCTTACACTCCGAG 150010  
 QY 2389 ----- 2388  
 Db 150009 TAGCTGGGCCAT 149950  
 QY 2389 ----- 2388  
 Db 149949 CTTGGCCAGGCTGTCTTTGAACCTCTGATCTGTGATCTACCCGCTTGGGCTCCCAAG 149890  
 QY 2389 -----gacactataaaatttcatitga 2413  
 |||||

polyA\_signal

BASE COUNT	817 a	778 c	749 g	713 t	2 others
ORIGIN					

Query Match	42.28;	Score 1326.6;	DB 74;	Length 3059;
Best Local Similarity	85.38;	Pred. No. 1.2e-297;		
Matches 1548;	Conservative 0;	Mismatches 239;	Indels 28;	Gaps 5

[illegible]

Db	786	GTCA	CCCCACTCGTAAC	CTACAG	CTGCCCCC	CACTATGG	CCAAAGG	ATGATCT	TATTA	845															
Qy	910	gcc	acttggcc	cagatcc	cagcag	caaaaa	aggaga	aggagcc	agatcac	gcctctca	969														
Db	846	GCAG	ACTGTCAC	ACTATCC	ACAGGCA	AAAAA	AGAGAG	AGCGG	AGTAC	ATGCTCTTA	905														
Qy	970	cag	agcag	gccc	ccgcgc	ccgca	gggagt	ttt	tgtc	aggtgna	aggttga	aaaccca	10299												
Db	906	CAGA	ACCGAG	CTCCTCC	ACGTCG	CAGGAG	AGTTTGTG	ATGACG	GTGAAG	GTGTGG	CACTCA	965													
Qy	1030	ctc	gcaagg	agga	cggcg	caacca	caaga	aagt	tgg	caag	cgcaat	cgacgc	cgga	ca	tcg	10899									
Db	966	CCGCA	GAAGAG	CAGG	GTACCA	ATRAGA	AGGTGG	CCAA	CGGCA	ATGCTGT	GAGAAC	ATGC	10255												
Qy	1090	tg	gaatc	ctt	tg	ttc	aaag	tc	cccg	ca	g	cg	gca	ccca	ccca	ccg	ca	ctc	aa	g	tc	aa	g	11499	
Db	1026	TGGAG	ATCTTG	GGTTCA	AAAGTC	CCCCAG	CGCAG	CTGCC	CAAG	CGAC	AGCTCA	ATATAG	10855												
Qy	1150	ag	aga	aag	aca	ccca	ta	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	12099	
Db	1086	AAGA	GAAAG	ACTCC	AGTAA	AAAGAAC	CAAG	AGATG	GAAG	AAATAC	ATT	TTT	TGA	ACCTA	11455										
Qy	1210	g	ctc	t	g	g	g	g	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	12699	
Db	1146	GCC	TTGGG	GGATGA	AAATG	GAACCA	GTATRA	TAAG	AGATAG	TTTCAG	ATGCTT	CTTA	12055												
Qy	1270	g	t	c	a	t	c	a	g	c	a	g	c	t	g	c	t	g	c	a	g	c	t	g	13299
Db	1206	GCA	ATCAG	CAC	ACTCC	CAGCG	GGGAT	TC	CCCAT	TG	TG	CG	CAG	GGT	TG	CCAG	CTG	CG	12655						
Qy	1330	g	a	g	t	a	g	t	a	g	a	g	a	c	a	g	a	a	a	a	a	a	a	13899	
Db	1266	G	G	G	T	T	A	G	C	A	A	G	C	A	C	A	A	G	A	T	T	T	C	C	13255
Qy	1390	c	a	a	c	g	t	a	a	c	t	a	a	a	a	a	a	a	a	a	a	a	a	14499	
Db	1326	CA	CG	GTAC	TG	CCATAT	ATAG	CCCG	GAAT	TG	TG	TAC	G	GG	GG	CA	CTG	CC	CA	CA	CG	13855			
Qy	1450	a	g	a	c	a	c	t	t	t	a	a	a	a	a	a	a	a	a	a	a	a	a	15099	
Db	1386	AG	AC	AT	T	T	T	A	A	G	A	G	T	A	C	A	T	C	T	T	C	A	G	14455	
Qy	1510	c	c	t	c	t	g	a	c	a	a	c	t	c	t	c	t	c	a	a	a	a	a	15655	
Db	1446	C	C	T	T	G	A	C	A	C	T	G	A	C	T	T	T	C	A	G	A	C	C	15055	
Qy	1570	t	c	c	c	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	16299		
Db	1506	T	C	C	C	A	G	A	C	A	C	A	A	A	G	A	G	A	G	T	G	T	A	15655	
Qy	1630	t	g	a	t	c	a	g	c	a	t	c	a	a	a	a	a	a	a	a	a	a	a	16699	
Db	1566	T	C	G	T	C	A	G	C	A	T	G	C	A	G	A	G	A	G	T	G	A	G	16255	
Qy	1690	t	c	t	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	17499		
Db	1626	T	C	T	T	A	A	A	C	T	C	T	T	G	A	C	T	T	G	A	C	C	A	16695	
Qy	1750	g	a	c	a	a	t	c	t	g	t	g	g	a	a	g	t	g	c	a	a	c	t	18099	
Db	1666	G	A	C	C	A	G	T	T	C	A	G	G	A	G	G	T	G	C						

## RESULT 11

OY	328	cttactctcgtgattgcagtcaccactaaactacaatgagagagtgcttaccoccca	387
Db	246	CCCACTCTCGGATGACAGTCACCACTACAGCTATGAGCATGCGTGGAGTGCCTATCCCCCA	305
OY	388	gtaactttaccattttccagtttccactttacttaltcaagtgaaactttctgtggag	447
Db	306	GATACCTTTTCCCAATTTCCAGTTTCCACCGTTACTCTTACCAATTTGAGTCTCCGGGAG	365
OY	448	gacacgaatttaatgcaagaagaaacaagaacgaagcttcgaaacagatctcgtgca	507
Db	366	GACAGCAAGTTTAATGGGAAGGAATAATAGACACACTCTGAAACATATATGCACTGCC	425
OY	508	aagcgtttgagatccctgcagaaatgagccctgcagagagagctgtgagtgaaatgaaag	567
Db	426	GAGCCCTGAGAGACTCTGCAGAGTGAACCCCTTACAGAGAGCTTAATGGAAAG	485
OY	568	aatccgaagaagaanaatctcaataaactlgaataatgaatgaagtgtttgagattgactta	627
Db	486	AATCAGAAGAGAAACCTTAATTAATTCAGAAATTAAGCCAACTGTTTGAATGTGGCTGA	545
OY	628	aacgaaccttgcctgtgaatttcgaaatgagtcgcgagagatggtgccaccccaatgaag	687
Db	546	AGCGGAATTTTGCTCTGTGAATTTTGAAGTGTGGCCCGGAGAGTGGCCGCACACATGAAGA	605
OY	688	acttltgaccaaagtttcggtltggygagtttgtggyggaagtgtgaagyggaagaagcaag	747
Db	606	ACTTTGTGACCAAGGTTTTCAGTTGGGAGATTGTTAGGGGAAGGAAGGAAGAAACAGA	665
OY	748	agatttcaagaanaaatgcccgcataatgcgtgttcttgagagagctgaagaagttaacggccc	807
Db	666	AGATCTCCAAAGAAAGAAATGCAGAGGCGCTCTGAGACAGCTTAAGGAGGCGTCCACCTC	725
OY	808	tgctctcagtttgaacgaagtaaaccttagatatcaaaagaanaacaacccaatgtaagc	867
Db	726	TACTCTGCTGTGAGACGAGTAACCCACAGATCAAGAGAAAGTCAAGCCACTCTCAAGC	785
OY	868	cacagacaagcccaagaatlatgcccaggygataatccgaattagccagacttggcccagatcc	927
Db	786	TACAGACTGCGCCCGACTATGCGCCAAAGGAGTGAATCTATTAGCAGACTGGCAGATCC	845
OY	928	agcagcgcaaaaaagagagagagccagagttacaagcttctacagaagcgagagctccgc	987
Db	846	AGCAGGCAAAAAAGAAAGAAAGAGACCGGAGTACATGCTCTTACAAAGAGAGTCTTCCAC	905
OY	988	gcgcagagagattgtgtatgacgtgaaatgttgaaacacacctgcagaagaaagcgga	1047
Db	906	GTCGCAAGGACTTTGTATCATCAGGTGAAGTTTGGCATACACCCACAGAGAGCAGGTA	965
OY	1048	ccacaagaagaatgtgccaagcgcaatgcagccgagagacatgcttgagataccttgtttca	1107
Db	966	CCAAATTAAGAAAGTGGCCACAGCGCAATGTGCTGAGAACTGTGGAGATCTCGGGGTCCA	1025
OY	1108	aagtcgcgcagcggcagcccaacaacccgcactcaagtcatgagagagaagacacccaata	1167
Db	1026	AAGTCCCCCAAGGCGCACCTCCGCAAGCGACACTCAAAATCAAGAAAGAACTCCAGTAA	1085
OY	1168	agaacaaggagatgagaaagaagttaaccttttgaacctgtgctctgtggatgaaagt	1227
Db	1086	AGAAACACGAGATGGAAGGAAGATTAACATTTTGTGAACCTTAGCCCTGGGATGTGAATG	1145
OY	1228	ggaactagtaataaagagatgagttcagaatgcttatctaagatcatcagacagctgacct	1287
Db	1146	GAAACAGTAATTAAGAGAGATGAGTTTCAGATGCGATGCCATTTAGCCATCAGACGCTGCAG	1205
OY	1288	ctggaattcttcccaatggtgcgcgaagtgcgcgaagctgtlagagttlgtacaagacatc	1347
Db	1206	CCGGATCTTCCCATAGTGTGCCAAGAGTTGCCACAGGCTGTGGGGTATATGTCAAAGACAC	1265
OY	1348	aacaccaagaatttaccagagcgactccgaatctctgcaagcgcaacgtaactgcacatga	1407
Db	1266	ATACCAAGAATTTTGCACGGGACAGCTTCCAATCTTCCCAAGGCACAGGTAACTGTCCATGA	1325

QY	1408	tagcccgagagtgctgtgtatatyggggacactcgtccacagccgagacatltttaaagaata	1467
Db	1326	TAGCCCGAGAAATGTTGTGTAACGGGGACACTCCGCCACAGCCGACATTTTAAAGACTA	1385
QY	1468	acatctcttcagagccacgtaccacatggacacctcaagagacccctctagcaactgagact	1527
Db	1386	ACATCTCTTTCAGGACATGTACCCCATGTGACCTTCGCACAAAGACCTCTGTAGCACTGTACT	1445
QY	1528	atctctccagagctccagagatltccaggttgtaataacaaagacttccccaacaaacaaaga	1587
Db	1446	ACCTTTCAGAGACCCAGGAGATTCCAGGTTGATTACAAAGACTTTCCCAAGAACACAAAGA	1505
QY	1588	acgaattgtatctcttatcaatctgctcctctcaagccacaccttgatcagcaatgtagtcg	1647
Db	1506	ACGAATGTGTATCTCTCATCAACCTCTCTTCACACCACCTTCGTGCAGCCATGGCATCG	1565
QY	1648	gcaagagatgtgagtccttcgacatgatatagctgctgagctgaacatcttaagttgctgctg	1707
Db	1566	GCAAGGATGTGGAGTCCCTGTCTCATGTATGTGCTGCACCTGACACTGAACTTTAAAGCTGCTGTG	1625
QY	1708	agttgaccacaagaaglacagagatgccaagatgccaagaacaggaacgagacaaatgtctgtg	1767
Db	1626	ACTTGGACCAACAGACAGACAGAGATGGCCAAAGACAGAAACGACACAGTTTCAGCGTGGC	1685
QY	1768	ggaagtgctgaacaccttctctgctgccaatgaacacattataaattcccaatatatactgaa	1827
Db	1686	GGAGGTGCTGAACCTATCTTGCGCATGAAACCTATTATATAA-CCCAACATATTTACCGAAA	1744
QY	1828	atactcga-aactgctcttgaacatttggaattctctgataccctcaatgagcgagaga	1883
Db	1745	ATACTGAGACTGCTTTTGAAATTTTGGATTCTGTATACCTCAGTGGGCCAAGACA	1801
RESULT	10		
LOCUS	AF227200	3059 bp	09-AUG-2000
DEFINITION	Rattus norvegicus stauflen isoform Stau-I6 (Stau) mRNA, complete cds, alternatively spliced.		
ACCESSION	AF227200		
VERSION	AF227200.1	GI:9754865	
KEYWORDS			
SOURCE	Norway rat.		
ORGANISM	Rattus norvegicus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
REFERENCE	1 (bases 1 to 3059)		
AUTHORS	Monshausen, M., Putz, U., Rehbein, M., Schweizer, M., Desroselliers, L., Kuhl, D., Richter, D. and Kindler, S.		
TITLE	Two rat brain stauflen isoforms differentially bind RNA		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 3059)		
AUTHORS	Monshausen, M. and Kindler, S.		
TITLE	Direct Submission		
JOURNAL	Submitted (20-JAN-2000) University of Hamburg, Institute for Cell Biochemistry and Clinical Neurobiology, Martinistrasse 52, Hamburg 20246, Germany		
FEATURES	Location/Qualifiers		
source	1..3059		
	/organism="Rattus norvegicus"		
	/strain="Sprague-Dawley"		
	/db_xref="taxon:10116"		
	/tissue_type="brain"		
	/dev_stage="adult"		
	1..3059		
	/gene="Stau"		
	227..1714		
	/note="RNA-binding protein; similar to Drosophila stauflen; contains four RNA-binding domains; rStau-I6"		
	/codon_start=1		
	/product="stauflen isoform Stau-I6"		
	/protein_id="AAI98119.1"		
gene			
CDS			

Dd	855	TCAGAGAGAAAAGTCAACCCCACTGCA-----AGCACGCCCGGATTATGGCCAAAGGA	908
Qy	898	lcaalccgcatltagccgaactlgycccaagatccagcagagcaaaaaaggagaagagccagagt	957
Dd	909	TGAATCCTATTAGTAGACTTGGCAACAGATCCAGCAGCCAAAAGGAGAAAGGAGCCAGAGT	968
Qy	958	aaagctccctcccaagagcagaagccctccgcgcgcgaaggagtltygtatgcagtlgaag	1017
Dd	969	ACATGCTCTCTTACGAACGAGAGTCTTCCAGTCCGCGGAGATTTTGATGCAAGGTAAAGG	1028
Qy	1018	tltygaaacccaacacgcgcgaaggaagagcgacccaaagaaggltgcccagcgcaatlgag	1077
Dd	1029	TTGGGCATCACACTCGAAGAGAGTGAGTACATATAGAAAGGTGGCCAGACCTATATGCTG	1088
Qy	1078	ccgagaacatlgctlgyaatactctlgytlccaaagtcccgacgcgtagccccaacaccg	1137
Dd	1089	CTGAGAAACATGCTGTGACATCTCTGGGGTTCAAAATTTCCAGGCGCAGCCTGCCAAGCCAG	1148
Qy	1138	cactcaagtlcaagagagagaagacacccaataagaaacccagggatgygaagaaaataacct	1197
Dd	1149	CACTCAAATCAGAGAGAAAGAACTCCAGTAAAGAAACAGAGAACCGGAAGAAATGAAGT	1208
Qy	1198	ttttttgaacctgcctcggggatggaataaaggagcctgtatataaagagatgaattcaga	1257
Dd	1209	TTTTTGAACCTTAGGCGCTGGGGAGTGAATAAGAACTGTATACAAAGGACGAGGAGTTCAGGA	1268
Qy	1258	lgyctatctcaagtcacacgaagcactgcctcgtlgyaatctctcccaatggtgccagagctg	1317
Dd	1269	TGCCTTATCTTAGGCATTCACAGACTGCGACAGCTGGAAATTTCTCCCATGGTGGCCGAAAGTTG	1328
Qy	1318	ccccagctgttagagtagtcaagagacatcacccaagaatttaaccagggcagctccga	1377
Dd	1329	CCGAGGCTGTGGGGTTTAGTCAAGAGACACACACCAAAATTTCCACAGGCGACGCTCCAA	1388
Qy	1378	atcttcgccaagggccaagctacgctacgtatagccgaagatgttltgtataggggcacct	1437
Dd	1389	ATCTCGCCAAAGCAACGGTATCTGCCATGTATACCCGAAGTTGTGTATCGGGGGCACTT	1448
Qy	1438	cgccaacagccgaagaccattttaagaataacatctcttcagggcaagtlacccaatgagc	1497
Dd	1449	CGCCCAACGCGGAGACCATTTTAAAGAGTATACATCTTTCAGGCGCACGTACCCCATGGAC	1508
Qy	1498	ctctcaagagacccctctgagcaactlgyactatctctccagaatlccaggaatlccagtlg	1557
Dd	1509	CTCGCACTAGACCTCTGAGCACTGTACTACTTTCCAGAGCCCAAGGGATTCAGAGTTG	1568
Qy	1558	aatacaagactlccccaataaacaagaagaattgtatctcttaaatgtctct	1617
Dd	1569	AATACAAAGATTTTCCCAACAAACAAAGAACAGAGTGTATCTCTCAATCAACTGCTCTT	1628
Qy	1618	ctagagcaacctctgaacagcatlgtatctgcaagaaatgtgaagtcctgcacatgaatag	1677
Dd	1629	CACAGCGCCTCTGTGATGATCATGGCATGCGCAAGGATGTGAAGTCTGTCTATGATATAGG	1688
Qy	1678	ctgagctlgyaacatctltlaaagtltgctlctlgyagtltgagaccaaaagtaagaatlgacca	1737
Dd	1689	CTGCACATGAAATTTTAAACTGCTGCTGTAGTGTGGACCAACAGAGCACAGAGATGCCAA	1748
Qy	1738	gaacaggaagaacgaccaaigtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1797
Dd	1749	GAAACAGGAAATGAGCCAGTTTACGCGTGGCGGAGGTGTGTGAACCTTTTCTGGCCACAAGT	1808
Qy	1798	catataaaatcccaacatatatactagaaataactbaa-aactgctltgaaaatltygaat	1856
Dd	1809	CATTATTAATA-CCCAACATATATATCTGAATAATCTGAAGACTGCTTGAATAATTTTGAT	1867
Qy	1857	tlctgataactccagltggtgcgaagata	1883
Dd	1868	ATCTGATATACCTCAGTGGGCCAAGACAA	1894

RESULT 9  
AF290989

LOCUS	AF290989	3041 bp	mRNA	ROD	09-AUG-2000
DEFINITION	Rattus norvegicus staufen isoform Stau-16 (Stau) mRNA, complete cds, alternatively spliced.				
ACCESSION	AF290989				
VERSION	AF290989.1	GI:9755162			
KEYWORDS					
SOURCE					
ORGANISM	Norway rat. Rattus norvegicus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;				
REFERENCE	Ratus. 1 (bases 1 to 3041) Monshausen,M., Putz,U., Rehbein,M., Schweizer,M., Desgroselliers,L., Kuhl,D., Richter,D. and Kindler,S. Two rat brain staufen isoforms differentially bind RNA unpublished 2 (bases 1 to 3041) Monshausen,M. and Kindler,S. Direct Submission Submitted (28-JUL-2000) Institute for Cell Biochemistry and Clinical Neurobiology, University of Hamburg, University Hospital Ependorff, Sudefeldstr. 24, Hamburg D-22529, Germany				
FEATURES	location/Qualifiers 1..3041 /organism="Rattus norvegicus" /strain="Sprague-Dawley" /db_xref="taxon:10116" /tissue_type="brain" /dex_stage="adult" 1..3041 /gene="Stau" /gene="Stau" 227..1696 /note="RNA-binding protein; similar to Drosophila staufen; contains four RNA-binding domains; rStau-16" /codon_start=1 /product="staufen isoform Stau-16" /protein_id="AAF98177.1" /db_xref="GI:9755163" /translation="MYKPVDPHSMOSTYSYGMKRGAVPRYPPEPPVPLLYOVL VGQDFNGKGMKRPVYKHDATPAHRTLSQEPRLPELVNGRESEENLNKSPISQV EIALKRLPVNFEVARESGPPHMLNFTFRTSVGEFVEGEGSKKISKNAARAVLEO LRLRLPRAVERVYKRLIKKSQPTCKLQTPADYGGQWNPISLQAIOQAKKEPEYK LIERGLRRREEVYQVGHHTAEAGTAKKYAKNAENMLEIGFVYQKQAPK ALKSEETPVKKPKDGGKRVFEPESPDEGNTSNKEDEFPMPYLSHOQLPAGLIPV EVAAGVSGCHHKDKFARAAPAPAKTATYAMTARELLVGGTSPATETLKSSISGH VPHGPTRPSPOLYLYLSRAQGFQVEYKDFPKNNKNECVSLINSSQPLVSHGICNDV ESCHDMAALNLIKLSLDDOSTEMPTGNGPVSAAGRC" 3025..3030 polyA_signal /gene="Stau"				
BASE COUNT	814 a	771 c	747 g	707 t	2 others
ORIGIN					
Query Match	43.1%;	Score 1354.6;	DB 74;	Length 3041;	
Best Local Similarity	86.1%;	Pred. No. 0;			
Matches 1548;	Conservative	0;	Mismatches 239;	Indels 10;	Gaps 4;
QY	89	cctccccccccgcgcgcgcgcgcgcctctccacgacgcactccgcctctctcc	148		
DB	13	CTTTCAGACCCCGCGCCGCGCGCCGCGCCCTTTCGAGACACCTCCGTCCTTCC	72		
QY	149	cttgctccctctctctctctctctctctctctctctcccccctccgcgcgcgcgc	208		
DB	73	CCGGGTCCTTCTTCTTCTTGCCCTCTCTTTTTCCTT-----TTCTTCTTGCCGCGCC	125		
QY	209	cagagaccgc	267		
DB	126	CAATACCGCCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	185		
QY	268	tagaacataatgcactgtcatgaaacttgaaaaaaaccaaagtcatagcctgttacc	327		
DB	186	TTGACCTCAATGCTGCTGCAACGGAAGAAACCAATGTATAAGCCCTGAC	245		

[illegible][illegible]

Db	2577	GATCCTCATTGTAGTCACATATTAACCTGCAGAGTGTATTTTCCAGAGTGTGCTGCATACTGTGA	2636
Oy	2890	attacatctccatgaaggctgaaaaaagaatgaacctgaagctttctgatatcaagctgtgtgct	2949
Db	2637	ATTACATCTCCATTTAGGGCTGAAAAAAGAAATGACCTTCTGTATACAGCTGTGTGCT	2696
Oy	2950	tttgaattgtgttaacbtatcacagaagaatggtgtgacacgaagcctgtgtgtccgta	3009
Db	2637	TTTGATGTGTGTACTGTATACACAGAAATGTGTGACATGAGAGCTGTGTGTGTGCTGA	2756
Oy	3010	tggaaacacctgtagcccttcgagatgaagcttaagctcttcacatcatctttacagctggaaat	3069
Db	2757	TGGAAGAGCTGGTAGCCCTGGCAGAGTTAGTACTGCTTCCATTCATTTGTTACGCTGGAGAT	2816
Oy	3070	ttttctcccatggaagtgaaagtaaacctaagtggtttgtcatcaataaatgtaatact	3129
Db	2817	TTTTCTCCCATGTGAAATGATTAAGTAAACTTATAGTTTGTTCATCAATTAAGTAAACT	2876
Oy	3130	aaa	3132
Db	2877	AAA	2879

RESULT 7  
AL133174/c

LOCUS	AL133174	113836 bp	DNA	PRI	14-JUN-2000
DEFINITION	Human DNA sequence from clone RP3-470L14 on chromosome 20, complete sequence.				
ACCESSION	AL133174				
VERSION	AL133174.15	GI:8573761			
KEYWORDS	HNC.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 113836)				
JOURNAL	Ramsay, H.				
COMMENT	Direct Submission Submitted (14-JUN-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Requests: clonerequest@sanger.ac.uk On Jun 20, 2000 this sequence version replaced gi:8388434. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at http://www.sanger.ac.uk/chr20 RP3-470L14 is from the library RPI-3 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2 IMPORTANT: This sequence is not the entire insert of clone RP3-470L14. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true right end of clone RP3-470L14 is at 113836 in this sequence. The true right end of clone RPI-15566 is at 100 in this sequence.				

FEATURES					
source	Location/Qualifiers				
	1..11836 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="20" /clone="RP3-47OL14" /clone_id="RPCI-3" 3061..3077 /note="Weak data" 95563..95627 /note="Single clone region. pUC subcloned to give 12x coverage. Assembly confirmed by restriction digest data."				
BASE COUNT	33974 a 25115 c 24233 g 30514 t				
ORIGIN					
Query Match	43.2%; Score 1357.4; DB 34; Length 11836;				
Best Local Similarity	99.9%; Pred. No. 0;				
Matches 1369:	Conservative 0; Mismatches 1; Indels 1; Gaps 1				
OY	1763 gtgtgaggatgcctgaacctttcttgccatagaaccatatataatcccacatatatc	1822			
Dd	34965 GTGTGGAGAGTCTGAACCTTTTCGGCATGACCATTATTAATCCCAATATATAC	34906			
OY	1823 tgaanaacttgaacttcgtcttgaanaatttggaaatttcgatattcccatgtggccgaag	1882			
Dd	34905 TGAAAAATCTGAAACTGCTTTAATAATTGGAAATTTCTGATACCTCCAGTGCGCGAG	34846			
OY	1883 acacagctggtaagaagatgtggcgacgacaaggaaacacagaacacaaggaagcg	1942			
Dd	34845 ACACGGTGGGTAAAGGATTTGGGCACACAGGGAAGACAAGAAMACMAAGAGCGG	34786			
OY	1943 ctgttgcc-ggcttgacttgccttggttgtlgtlgtatgtgcactcgtlgaccttgcgt	2001			
Dd	34785 CTGTGGCGCGCTGACTGTGCTGGGGTTTGTGTGATGGCACCTCGTGACCTGGCGGT	34726			
OY	2002 ccctcgcgaatagcagcttcgctgttgggaagaaggctctgccagcagtgttctccg	2061			
Dd	34725 CCTCAGCAATGACAGCTCCGCTGGGGGAAGAGGCTGCCAGCGGTCTCCCG	34666			
OY	2062 ggacaacagacatcacacaccttggcaacctcgtgtgtgtcttctttccccctgt	2121			
Dd	34665 GGACACACAGATCCACACCCTGGGCACTCCGTGTTGATTTTTTCCCCTGTGT	34606			
OY	2122 gaagaagaacgycagcagcccccttcacgttgcctcacctcagacacatcgggacaac	2181			
Dd	34605 GAAGAAGAAAGCGACGACGCCCTTCTCAAGCTGGCTACTCAGACACATTGGGACAAAC	34546			
OY	2182 ccttgacagccatgcacagagaagaagcctttgacccgccccagaagctaaaagacccaaga	2241			
Dd	34545 CCTGGAACACCCATGCGCAGGAAGAGGCTTTGACCGCCCCCAAGAGCTAAAGACCCGAA	34486			
OY	2242 aaataaatgtcttcactcactcagcgtgaaccaactttctagtgtgcacagccccaca	2301			
Dd	34485 AATATAATGCTTCTTACTACAGCTGACCCAATTCTTAGTGTGCCAGGCCCCACAC	34426			
OY	2302 ctccctcagtlacccacacacatcacacactgtcttctcttccaaacagtatctlatctta	2361			
Dd	34425 CTCCTGAGTACCAACACATCACCACCTGCTTCTTCCAACAGATGATGATTTCTTA	34366			
OY	2362 attcataattcttcttgtatgtatgatgacactataaatttcattttgaatttct	2421			
Dd	34365 GTTTCATTATTTTCTTTGATGATATGACACTATATATAAAATTTTCATTGGAAATTTCT	34306			
OY	2422 caatgtatctagttaaataagacagtttggaaacttgcctaagcttgactttaacaata	2481			
Dd	34305 CAATTGTATCTAGTTAAATAGACAGTTTGGAAACTTGCTCAGACGTGACTTATCAAAA	34246			
OY	2482 atctaacgacaagaatcatcatcatgttatgtgttgagacatttcttcatattgact	2541			
Dd	34245 ATCTAACGACAAATCATATCCATGTGTATGTGTGATGAGATTTTATTTTCATTGTA	34186			
OY	2542 aaccacgagcagtttcagltgatgcaaatgtgtgcaccttgttctcagctgaaacagtcct	2601			



QY 673 caccacacatgaagaactcttgtagccaaaggttccgtctggaggagttctgtggggaagtg 732  
 |||||  
 Db 421 CACCCACATGAGAACTTTGTGTGACCAAGGTTTGCGTGGGAGTTGTGGGGAGAGTG 480  
 QY 733 aagggaagaagaagaatttcaaagaataatgcgcacatagctgtctcttgaagagctga 792  
 |||||  
 Db 481 AAGGGAAGACCAAGAAATTTCAAGAAATATGCGCATAGCTGTCTTGAGAGGCTGA 540  
 QY 793 agaagttacccctctccctgcagttgaacagatgaagaaagtaagaataaagaacaa 852  
 |||||  
 Db 541 AGAAGTTACCGCCCTCTGCACTTGAACGAGTAAAGCTAGATCAAAAAGAAAAA 600  
 QY 853 aaccatagtcacagacacagaccagaataatggccagggagatcaatccgaatga 912  
 |||||  
 Db 601 AACCCATAGTCAAGCCACAGCAAGCCAGATATGGCCAGGGATCATATCCGATTAGCC 660  
 QY 913 gactgcccagatccagcagggcaaaaaaggaaaggagccagagttacagctctccacag 972  
 |||||  
 Db 661 GACTGCGCCAGATCCACAGCAAGAAAAAGAGAGAGCCAGATTACAGCTCTCCACAG 720  
 QY 973 agcgaagccctcccgccgcagcagagagtttgtagcaggtggaaggttggaaccacac 1032  
 |||||  
 Db 721 AGCGAGGCTCTCCGCGCGAGGGAGTTTGATGACAGGTGAAGGTTGGAACACACACTG 780  
 QY 1033 cagaagaacagcagcacaagaagaagtgccaaagcgaatgcagccagagacatgctg 1092  
 |||||  
 Db 781 CAGAGAGAACGGGACACACAGAAAGGTGGCCAAAGCCCAATGCAGCCGAGAACACTGCTGG 840  
 QY 1093 agatccttggtttcaaagttcccgccagcgcagccacacaaacccgcactcaagtcagag 1152  
 |||||  
 Db 841 AGATCTTTGGTTTCAAAAGTCCCGCAGCG--CAGCCACCAAAACCCGCACTCAAGTCAGAGG 898  
 QY 1153 agaagaacaccataaagaacacagggatggaagaagaatgaac--tttttgaacctggc 1211  
 |||||  
 Db 899 AGAGAACACCCATTAAGAAACAGGAGTGAAGAAAGTAACTTTTGTGAACCTGGC 958  
 QY 1212 tctggagat--gaaaatggagtagtaataaagagatgaagtcagagtccttaactaa 1270  
 |||||  
 Db 959 TCTGGGATGGAATAATGGAGTACTAATAAAGAGATGAGTTCAGATGCTTATCTAAG 1018  
 QY 1271 tcatcagcagctgcctgcctggaattcttcccaatggcccgaggttcgcccagagctgtaa 1330  
 |||||  
 Db 1019 TCATCAGCAGCTGCTCTGTGAATTTCTCCATGCTGCGCCAGGCTCGAGCTGTAGG 1078  
 QY 1331 agttagtcaagagacataccaccaagaatttaccagggcagctccgaatcctgcacagc 1390  
 |||||  
 Db 1079 AGTTAGTCAAGAGACATCACACAAAGATTTTACAGAGGACCTCGAATCTGCGCAAGGC 1138  
 QY 1391 caggttaacatccaatgaagccagagttgtatgggggacacccctgcacacagccga 1450  
 |||||  
 Db 1139 CAGGTAACCTGCCATGATAGCCCGAGCTTGTGTATGGGGGACCTGCGCCACAGCCGA 1198  
 QY 1451 gaccattttaaagaataacatctctcaagccacgtlaacccaatgaacctctacagagac 1510  
 |||||  
 Db 1199 GACCATTTTAAAGATTAACATCTCTCAGGCGACGTACCCCATGAGACTCTCACAGAGAC 1258  
 QY 1511 ctctgagcagctgacatctcttccagagttcagagatccaggttgaatacaagactt 1570  
 |||||  
 Db 1259 CTCTGAGCACTGACTATCTTCCAGAGTCCAGGAGATTCAGAGTTGAATCAAGAGATT 1318  
 QY 1571 ccccaaaaaacacagaacgaattgtatctctatcaatgcctctctcagcacactt 1630  
 |||||  
 Db 1319 CCCCCAAAACACAGAAAGAAATTTGTATCTCTATCAATTGCGCTCTCAGCCACT 1378  
 QY 1631 gatcagcagatgatcagcagaagatggaggtccctgcacatgaatgctgcgtgaacat 1690  
 |||||  
 Db 1379 GATAGGCATGATCGCAAGATGTGAGTCTGCGCATATATGCTGCGCGAATAT 1438  
 QY 1691 cttaaagttgctctgaggtggagccaaagaatfacagagatggcacaagaaggaagag 1750  
 |||||  
 Db 1439 CTTAAAGTGTCTGTGATGAGTTCAGCAACAGATACAGATGCGCAAGAGAAAGGG 1498  
 QY 1751 accaatgtctgtgtgtggaggtgctgaaccttcttgccatgaacacataataatcc 1810

Db 1499 ACCAATGCTGTGTGGGAGGTGCTGAACCTTTCTGGCCATGAACATATATAATCC 1558  
 QY 1811 caacataatactgaataactgcttgaataatttgaattcttgatacccca 1870  
 |||||  
 Db 1559 CAACTATATACGAAAAATACGTGAACCTGCTTTGAAAAATTTGGAAATTTCTGATACCTCA 1618  
 QY 1871 gtaggcagagacagcgttggtaaagaatgtggcagcagcaagaagaagaagaac 1930  
 |||||  
 Db 1619 GTGGCCGAGAGACACGCTGGGTAAAGATGTGGGCAAGCAGAGGAAGACACAGAAAC 1678  
 QY 1931 acaagagagcgtctgtgccc--gactgagcgtgtggtgggttctgtgataagccactcgg 1989  
 |||||  
 Db 1679 ACAAGGAGCGCGCTGTGGCGCGCTGACGTGGCGGGGTTTGTGTGATGGCCACTCGG 1738  
 QY 1990 tgaccctgcggtccctacagcaatagcagctgcctgttgggaagaagagctgcgcccagcc 2049  
 |||||  
 Db 1739 TGACCTGGCGGTCCCTACGCAATGACAGCTGCTGTGGGGAAGAGGGGCTGCCAGGCAG 1798  
 QY 2050 ctggttctcccgagacacacagatccacacccctgggcaacctcgtgttgcctttt 2109  
 |||||  
 Db 1799 CTGCTTCTCCGGGACACACAGATCCACACCTGGGACCTCGTGTGTGGTCTTTT 1858  
 QY 2110 ttcccccctgtgaaagaagaagaacgagcagcccttctcaagcttggtcactcagacac 2169  
 |||||  
 Db 1859 TTTCCCTGTGTGAAGAAGAAAGACGACGACCCCTCTCAAGCTGGGCTCACTCAGACAC 1918  
 QY 2170 attgagacaaacctgtgacagcaltgcagaagaagagccttgaacggcccaagactaa 2229  
 |||||  
 Db 1919 ATTGGAGCAAA--CCTGGACACCCATGCGCAGAGAGAG--CTTTGACCGGCCCCCAAGCTAA 1976  
 QY 2230 aagacacagagaataataaagcttccctactcagcgtgagacacacttcttagtgcga 2289  
 |||||  
 Db 1977 AAGCACAGAGAAATCAAAAGCTTCTCTACTCAGCGTGAACCACTTTCTAGTGTGCA 2036  
 QY 2290 cggcccccacacccctcctgcagttacccaacacatcacacactgtcttcttccaaacagtga 2349  
 |||||  
 Db 2037 CGGCCCAACACACCTCTGCACTACCCACACATCACCACCTCTTCTCTCAACAGTGA 2096  
 QY 2350 tctgtatcttagttcaatatttcttcttgaattgatatgaactataaatttcat 2409  
 |||||  
 Db 2097 TCTGTATTTCTAGTTTATATATTTCTTTTGTGATGATGACATATATAAATTTTCAT 2156  
 QY 2410 ttggaatttctcaattgtatctagttaaatagacagtttggaaactgtctgagactg 2469  
 |||||  
 Db 2157 TTGAATTTTCTCAATTTGTATCTTAAATAGCACAGTTTGGAACTTGTCTAGACTG 2216  
 QY 2470 acttlatcaatcaatcaacccgacaagaatcatatcatgtgtatgtgtttagacatttt 2529  
 |||||  
 Db 2217 ACTTTATCAATATATCTAAACGACAAAGATCATATCCATGTGTATGTGTAGACATTTT 2276  
 QY 2530 attcatctgaactaacccagagcagtttcagtgatgacaaatgtgtgcctctgtgtcagc 2589  
 |||||  
 Db 2277 ATTTCATTTGACTTAACCCAGGACAGTTTCAAGATGCAAAATGTGTGCTCTGTGTAGC 2336  
 QY 2590 tgaacagctctgagacttcaaaaaccttgataaagtctcccaagttgtataatgtga 2649  
 |||||  
 Db 2337 TGAACACGCTCTGACCTTTCAAAAAACCTTGAATAGTCTCCACAGTTGTATTAATTGGA 2396  
 QY 2650 caatttaggaattttaaactttagatgaltcatctgtgttcaatttatttattt 2709  
 |||||  
 Db 2397 CAATTTAGGAATTTTAAACCTTAGATGATCATTTGTGCTCAATTTTATTTTATTTAT 2456  
 QY 2710 ttgttaatgcaaacagagacttaaatgaacttgatctctgtttaaagattataaaaa 2769  
 |||||  
 Db 2457 TTTGTATATGCAAAACAGACTTAATGAATGAACTTGTATCTCTGTATTAAGATTATTAANA 2516  
 QY 2770 acattgtatctatacatatgactcttgagagacttagcttcaactacatcagatat 2829  
 |||||  
 Db 2517 ACATTGTATCTATACATATAGGCTCTTGAAGACTTAGCTTCACTACACAGAGATAT 2576  
 QY 2830 gatctcaatgtagtcatataaacaactgcaagagatttccagaggtgctgatactgtla 2889  
 |||||



KVSVEFVEGESEKSKISKKNNAALVLEELKPLPAVERKPRIKKKTIPIVKPO  
TSPEYGGINPISRLAIOJQAKREKEPPTLLTEBGLPREPFWOVKVGNTAGTGS  
TNKKYAKRNAANMELIIEFKYRPOPTKPKALSEKPIKRGKGRKRTFFERPSGD  
ENGTSNKEDEFRMPYLSHQDPLPMPEVAQAVGSGHHTDFTFAANPAKAT  
VYAMARLELYSGTSPRTAETILKNNISGHPGLTRSEOLDLSRVOGQVEYKD  
FPKNKNEFVSLINCSQPLISHGIGKDVESHMAALNLIKLSLDOOSTEMPRRT  
GNGPMSVCGRC\*

BASE COUNT 965 a 875 c 790 g 876 t

ORIGIN

Query Match 92.2%; Score 2896; DB 10; Length 3506;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2896; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 247 cagaagcataaccctactgttagaactaaatgcactgtgcatacttgaaaaaac 306  
|||||  
DB 611 CAGAAAGCATMAACCCCTACTGTGAACATAATGCACTGTGCATGAACCTTGAAAAAAC 670  
caatgtataagcctgttagccctactctgcgaatgcagttccacttaactataacataga 366  
|||||  
DB 671 CAATGATATAGCCTGTGTGACCTTACTCTGGATGCACTCCACCTAATACTACAAACATGA 730  
gaagaggtgcttatacccccgaagttactttaccatltccagttccaccttaactatc 426  
|||||  
DB 731 GAGGAGGTGCTTATCCCGAGGTACTTTTACCAATTTCCAGTTTCCACCTTTACTTTATTC 790  
aagtgaaacttctgttgaggagaacaaatttaaaggcaagaaagaacaaagcagctg 486  
|||||  
DB 791 AAGTGAACCTTCTGTGGCGAGACAAATTTATGCAAAAGCAAGACAGAGCTG 850  
cgaagacagatgtctgtccaaagcgtttagagatcctctgcgaatgagccctccagaga 546  
|||||  
DB 851 CGAAGACGATGTGTGTGCGCAAAAGCGTTGAGATCCTGCAATGAGCCCTCCAGAGA 910  
gactggaagtgaatggagaagaaatccgaagaagaaatcctaataatcgaataatc 606  
|||||  
DB 911 GGTGGAGGTGATGGAAGAGATCCGAACAATAATCATAATTAATCGAATAATAGTC 970  
aagtgattgagatgtcaactaaagaaactgtcctgtgaattccaggtgagcccgagaga 666  
|||||  
DB 971 AAGTGTGTTGATGTGACCTTAAACGGAACCTTGCTGATTTGAGAGTGGCCCGGAGA 1030  
gtggccacccacatagaagaacttgtgacaaagtttgcgttggagagtttggggg 726  
|||||  
DB 1031 GTGGCCACCCACATGAACAACTTTGTGACCAAGGTTTGTTGGGGAATTTGTGGGG 1090  
aaggtgaaggaagaagaagaatltcaagaanaaatgcgcacatagctgttcttgaag 786  
|||||  
DB 1091 AAGGTGAAGGGAAGAACGACGAAGATTTCAAGAAAAATGCCGCCATRAGCTGTTTCAGG 1150  
agctgaagaagtataccgcccctgtcctgtcagttgaacgagtataagcctagatacaaaaga 846  
|||||  
DB 1151 AGCTGAAGAAAGTTAACGCCCTCTGCTGCAATGTAACGAGTAAGGCCAGAAATCAAAAAAGA 1210  
aaacaaaaccatagttcaagcacaagaagaagccagaataatgagcagggatcaatcga 906  
|||||  
DB 1211 AAACAAAACCATAGTCAACCCACAGCAAGCCAGAAATGTGGCCAGGGGATCAATCCGA 1270  
tttagccactgagccagatccagcagaagaaagaagaagagagagagagataacgttc 966  
|||||  
DB 1271 TTAGCCGACTGGCCAGATCCAGAGCAAGAAAAAGGAAGAGCCGAGATACAGCTTC 1330  
tcacagaagcagcctcccgcgccgaggaagtttctgattcaggtgaaggttgaagacc 1026  
|||||  
DB 1331 TCACAGAGCAGAGGCTCCCGCGCGCAGGAGTTTGTGATGCAAGTGAAGTTGGAAGACC 1390  
aacctcagaagaagaagggacacaagaaggttgccaaagcgaatgacgcgagaga 1086  
|||||  
DB 1391 ACACGTGAGAGAGACGCGGACCAACAAAGAGGTGGCCAAAGCCAGCCGAGAGACA 1450  
tgcttgagatccctgtgttcaaaagtcgcgagggcagccacacaaaccgacactaagt 1146  
|||||

DB 1451 TGCTGGAGATCCTTGGTTTCAAGTCCCGGACGCGGACGCCACCAAAACCGCAGCTCAAGT 1510  
QY 1147 cagaagagaagaacaccataaagaacccaggggaatggaagaagaatgaacttlttgaac 1206  
|||||  
DB 1511 CAGAGAGAGAGAACCCCATTAAGAAAGAACCCAGGAGATGAAGAAAAAGTAACCTTTTGAAC 1570  
ctggtctgggggaatgaaaatgggactagtaalaagaaggaatgagttccagatgccttacc 1266  
|||||  
DB 1571 CTGGCTCTGGGGATGAAGAAATGGCATGATATTAAGAGATGATGATTCAGAGATCCTTATTC 1630  
taagtcatacagcagctgcctgcctggaattcttcccatggtgcccgaagttgcaccaagctg 1326  
|||||  
DB 1631 TAAGTCATCAGCAGTGCTGCTGATGATTCCTTCCATGCGGAGAGTCCGACAGGCTG 1690  
taggagttagtcagaagacatcacacaagaatlttaccagggcagctccgaatccgtgcca 1386  
|||||  
DB 1691 TAGAGTGTACTCAAGGACATCACACCAAAATTTACAGGCGCAGCTCCCAATCCGCA 1750  
aggcacaagtaactgcataatagccgagagttgtgtatgaggggcagctgcaccaag 1446  
|||||  
DB 1751 AGGCCACGTAACCTGCATATAGCCCGAGAGTTGTTGTTAGGAGGACCTCGCCACAG 1810  
ccgagacacatlttaagaataacatcctcctcaagcagctaccaccaatggaactctcaaga 1506  
|||||  
DB 1811 CCGAGACCATTTTAAAGAAATACATGCTCTTCAAGGCCAGTACCCCATGACCTTCACGA 1870  
gaacctctgagcaactggaactccttccagagttccaaggaatccaggttgaatacaag 1566  
|||||  
DB 1871 GACCTCTGTAGCAACCTGAGACTATCTTCCAGAGTCCAGGAGATTCAGAGTTGAATACAAAG 1930  
acttccccaacaaacaaagaagaacttgtatcctcttaccatctcctcccaagccac 1826  
|||||  
DB 1931 ACTTCCCAAAACCAACAGAACGAATTTGTATCTCTTATCAATTTGCTCTCAGACAC 1990  
ctctgaatcagcaatggtatcagcaaggaatgagagttcctgccaatgattgctgcctga 1686  
|||||  
DB 1991 CTCTGATACACCATGATGATGCGCAAGAGATGGAAGTCCCATGATATGCTGCGCTGA 2050  
acatcttaagttgtctgtcgtgagttggaacacaagaatcagagatgccaagaagga 1746  
|||||  
DB 2051 ACATCTTAAAGTTGCTGTGTGAGTTGAGTGCACCAAAAGTACAGAGATGCCAAGAACAGGA 2110  
acggaccaatgtctgtgtgtgaggaggtgtcgaaccttlttgcgaatgaacatataa 1806  
|||||  
DB 2111 ACGGACCAATGCTGTGTGTGGAGGTGCTGAACCTTTTGTGGCCATGAACCTTATAAA 2170  
atcccaacatataactgaataactgaactgtcttgaataatlttgaatlttgaatc 1866  
|||||  
DB 2171 ATCCCAACATATATACGAATAATACGTAACCTGTTGAATAATTTGGAATTTCTGATACC 2230  
tccagtgaggcagagagacacggtggtgaaagatgtggcagcagcaggaagaagaacag 1926  
|||||  
DB 2231 TCCAGTGGCGCAGAGACACAGGTGGTGAAGATGTGGCAGCAGCAGGAAAGAACAG 2290  
aaacacaaggaagcagctgtgtgcccgtgtgactgtgtgtgtgttlttgtatgagccact 1986  
|||||  
DB 1927 AAACACAAGAGGCGGCTGTGTGGCGCTGAGCTGTGTGGGTTTGTGTGATGGCCACT 2350  
cgtgtaccctggcgttccctacgcgaatagcagctgcctgttgagggaagaagagctgcaccagc 2046  
|||||  
DB 1987 CGGTGACCTGGCGGCTCCCTACGCAATRAGCAGCTGCTGTGGGAAGAAAGGCTGCCAGC 2410  
cagctgttctcccggaacacagcagatccacaacccctgggcaacctcgtgttgttctt 2106  
|||||  
DB 2411 CAGCTGTTCTCCCGGAGACACAGATACACACCTGTGGCAGCTCCGTTGTGGCTT 2470  
tttttccctgtgtgaaagaagaagcagcagaccccttccaagctgtgttaactcaga 2166  
|||||  
DB 2471 TTTTTCCTCTGTGTGAAGAAAGACGACACCCCTTCTAAGCTGTGCTACTCAGA 2530  
cacatgagacaacccctgacagcagcatgcccagagaagagccttgaacggcccgagagc 2226  
|||||  
DB 2531 CACATTGGCAAAACCTTGACAGCCATGCCACAGAGAGGCTTTTGAACGCGGCCAGAGC 2590

Db	1801	TTGATGCTCTTATCATTAATTCGTCCTTCACGACACTCTGTATCAAGCATGTFATCGGCAAG	1868
Oy	1654	atctgtagtcctccatcgatataatgctctgccttgaaacatcttaagttgcctgtctgaatttg	1713
Db	1861	ATGTGGAGTCTCTGCACATGATATGGTCGCGCTGACATCTTTAAAGTTGGCTGCTCGAGTTGG	1920
Oy	1714	acgaacaagtaacaggaatctgcaagaacaggaaacgaccaaatgctctgtgtgtgtgtgaat	1773
Db	1921	ACCAACAAGATACGAGATATCCCAAGAACGGAACCGACCAATATGCTGTGTGTGGAGCT	1980
Oy	1774	gctaaacctcttcygcacatgaacacatlaaatcccaacataatctgaaatactg	1833
Db	1981	GCTACACTTTTTCGGCGCATGAAACATATAAATCCCAACATATATACTGAAATTAATG	2040
Oy	1834	aaactgctcttgaaaaatttgaaattcttgabacttcagtcagtcgagagaacacgctggt	1893
Db	2041	AAACTGCTTTGAAAATTTGGAAATTTTGATATCTCCAGTGGCGCAGAGACACGCTGGT	2100
Oy	1884	aaagaatgttgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag	1953
Db	2101	AAAGATATGTGGCGACGACGAGGGAACAACAACAAACAAAGAGAGCGGCTGTGGCCGG	2160
Oy	1954	tgagactgtcctgggggtttgt	2013
Db	2161	TGGACTGTGCTGGGGTTTGT	2220
Oy	2014	gcaactgctctgttgaggagaagaagcctgcccagcagcagcagcagcagcagcagcagcag	2073
Db	2221	GCAGCTGCTGTGTGGGAAGAGGGCTGCCAGCACCTGTGTCCTCCGGAGACACGACAG	2280
Oy	2074	atccaacacctgggacacctgcgt	2133
Db	2281	ATCCACACCCCTGGGACACCTCGTTGGCTTTGGCTTTTCCCTGTGTGAAGAAAGAAC	2340
Oy	2134	ggacagacacctcttcaagcttgctcaactcaacacacatitgggaacaaaccttggagaacga	2193
Db	2341	GGCAGACACCTTTCTCAAGCTGGCTACCTCAGACACATTTGGACAAACCTGTGGACAGCA	2400
Oy	2194	tgccaagaagaagcctcttgacccgcccacagagctaaagaacacagaagaataatcaaatgct	2253
Db	2401	TGCCAGAGAGGGCTTTTGACCGGCCCAAGAGCTAAACACACAGAAATCAATATGCT	2460
Oy	2254	tctactcaagcgtgaaccaactttctctagtggtgcaagggccccaacctctcagctac	2313
Db	2461	TCTACTACAGGTGACCCCAACTTTTCTAGTGTGCCAGGCCACACACCTCTCTGCAGTAC	2520
Oy	2314	ccaaacatcaacacactgtctctctctccaaagtgatctgtatctctgtgtgtgtgtgtgtgt	2373
Db	2521	CCCAACCAATCAACACTGCTCTTCTCTTCCCAAGTGATCTGTATCTTGATTCATTATTT	2580
Oy	2374	tctcttgatgatatatgacatatataaataatlttcaatlttgagaatcttccaattgattacta	2433
Db	2581	TCTTTTGATTTGATATGACACTATATATAATTTTCATTGAGAAATTTTCCANTTGATCTA	2640
Oy	2434	gttaaatagcaagctttggaactgtgtcgtagaacttataatcaataatcaacagca	2493
Db	2641	GTTAAATAGCACACTTTGGAAACTTGTCTGAGACTGACTTTCAATATATCAACCGACA	2700
Oy	2494	aagaatcatcaacagtgatgtgtgtgtgaagcaatlttattatlttgaaatcaaccaggagag	2553
Db	2701	AAGATCATATCCATGTATGTATGTGTAGACATTTTATTTCAATGACTTCAACCGAGACG	2760
Oy	2554	tctcaagtgaatgaatgtgtgcctctgtgtcagcttganaaagctccctggaacttcaaaa	2613
Db	2761	TTTCAGATGAGCAAAATTTGTGTGCCCTGTGGTTCAGCTGAACAGTCTGGACTTTCAAA	2820
Oy	2614	accttgaaataagctcccaagcttgtataaatltgacaaatlttagaatlttaaaccttag	2673
Db	2821	ACCTTGATATAGTCTCCCAAGCTGTGTATTAATTTGACATTTTAGAAATTTTAACTTAG	2880
Oy	2674	atgtaaatgtgtgttcaattttatcttattttatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	2733
Db	2881	ATGATCATTTGTGGTCCATTTTATTTTCAATTTTATTTTGTAAATGCAACAGCGCTTAA	2940

QY	2724	atgaacttgcctcgtgtttaaagaatcaatcaaaacatgtagatctatacaatagc	2793
Db	2941	ATGAACCTTGTGATCTCTGTTTAAAGATTATTTAAACAAATGTGTATCTATACATATGCG	3000
QY	2794	ctctgaagacttagcttccaactacacacagagatatgatctccatgtagtccataaac	2853
Db	3001	TCTTGAGGACTTAGCTTTCACACTACACTACAGATATGATTCOCATGATGTCATTAAC	3060
QY	2854	ctgcagatgatcttccagagatgctcgatctgtaactgttaactcaactccatctaggcctgaa	2913
Db	3061	CTGCAGACTGATTTTCCAGAGAGTGCTCGATACGTGTAATTAACATCTCCATATGAGGCTGAAA	3120
QY	2914	agaagacctacgcttctcgtatacagacgcgtgtgtgtcttctgaagtggtgtctactgaaca	2973
Db	3121	AGAAAGACCTACGTTTCTGTGTATACAGCGTGTGTGCTTTGATGTGTGTACTGTACACA	3180
QY	2974	gaagctgtgcactgaaggctcctgcgtgtgtgtccgcatgtaaaacctgtagccctgcag	3033
Db	3181	GAAAGTGTGCACACTGAGGCTCTGCGCTGTGTGTCGATGGAACCTGGTAGCCCTGCAG	3240
QY	3034	ttaagtaactgctccatcgaatctgttctacgctggaatcttctcccatggaatgtaagta	3093
Db	3241	TTAAGTACTGCTCCATTCATCATGTTTAGCGTGGAATTTTCTCCCAAGGAATGTAACGA	3300
QY	3094	aaacttaagtgcttgatcgaataaataatgtaactataaaaaaanaa	3142
Db	3301	AAACTTAAGTGTGTGTCAATCATTAATGATGTAATAAAAAAAAAA	3349
RESULT	5		
AF061939			
LOCUS	AF061939	3506 bp	mRNA
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
FEATURES			
source			
gene			
CDS			



QY 1516 agcaactgagctatcttccagagatccaggatccaggttgaatatacaagacttcccca 1575  
 Db 1558 AGCAACTGGACTATCTTCCAGAGTCCAGGATTCAGAGTTGAATACAAAGACCTCCCA 1617  
 QY 1576 aaaaacaagaagaacttgatctcttaataatgtctctcccaaccctctgtca 1655  
 Db 1618 AAAACAACAAGAAGAAATGATCTCTTATCAATTCCTCTCCACCACTCTGTATCA 1677  
 QY 1636 gcaatggtatccgcaagatgtgagtcctcccaatgatatgctgcgtgaactctaa 1695  
 Db 1678 GCATGATTCGGCAAGAGATGTGAGTCCCATGATATGGCTGCGCTGAACATCTTA 1737  
 QY 1696 agtgcgtctgagttgagccaacaagaatcagagatgccaagaacaagaaacgaa 1755  
 Db 1738 AGTGTGCTGTGAGTTGAGCAACAAGATACAGATGCCAAGAACAGGAACGAGACCA 1797  
 QY 1756 tgcgtggtgagagtgctgaaccttctctgccaatgaatcttaataatccca 1815  
 Db 1798 TGCTGTGTGGAGGTGCTGAACCTTTTCTGGCCATGAAACATWAAAAATCCCAACA 1857  
 QY 1816 tataatacgaataatacgaactgcttgaaatltgaaatlttgatctccatccagtg 1875  
 Db 1858 TATATACTGAATAATACGAAGACTGCTTGAATAATTTGGAATTTCTGATACCTCAAGTGG 1917  
 QY 1876 ccgagagacacggtggtaagaatgltggcagcagcaggaagacacagaaacaaag 1935  
 Db 1918 CCGAGAGACACGGGTGAGTGAAGATGTGGCAGACAGAGGAAGAACAAAGAACACAG 1977  
 QY 1936 gaagcgagctgtgccc-ggcgtgagctgtgctggaggttctgtgtatgagcactgg 1994  
 Db 1978 GAGCGGCTGTGGCCGGGCTGGAGTGTGTGGGTTTGTGTATGAGCCACTGGTGACC 2037  
 QY 1995 tggcggtccctgaacgaatgaacgtgctgtgggaagaagggctgcacagcagctgt 2054  
 Db 2038 TGGCGGTCCCTAGGCAATACGACGTGCTGTGGGGAAGAAAGGCTGCCAGCAGCTGCT 2097  
 QY 2055 tctccggagacacagcagatccacacctgagcactccgtgttggcttcttctcc 2114  
 Db 2098 TCTCCGGGACACAGCAGATGCCACACCTTGGGACCTCCGTGTGTGTGTGTGTTC 2157  
 QY 2115 ccgtgtggaagaagaagcagcagccttctcaagctgctgctactcaagacatltg 2174  
 Db 2158 CCGTGTGAAAGAAAGAAACGGCAGACCCCTTCTCAAGCTGGCTCAGACACATTTG 2217  
 QY 2175 gaaacacccctgagcagcagcagagagagccttgaccggcccccagaggtcaaaag 2234  
 Db 2218 GACAAACCTCGACAGCAGCATGCCAGAGAGGCCCTTGGACGGCCCCAGAGCTPAAAGCA 2277  
 QY 2235 ccagagaaaatcaaatgctctactcaagcgtgacccaacttctctagtgccagggc 2294  
 Db 2278 CCGAGAAATCAAAATGCTTCTACTCAGGCTGACCCAATTTTCAATGTGACAGCCGC 2337  
 QY 2295 ccaacacccctgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2354  
 Db 2338 CCACCAACCTCTCAGACAGCAGCATGCCACTGCTTCTTCTCAACAGAGATCTGT 2397  
 QY 2355 atctgaatctatcttctcttgatgatatgacacacataaatttcaattttag 2414  
 Db 2398 ATCTTGAATCAATATTTCTTGTGATGATGACACATATPAAATTTTATTTGAG 2457  
 QY 2415 aattctcaatglatctagtaataatgacagtttgaaactgtctgagactgtctt 2474  
 Db 2458 AATTTCGATGTATCTATCTAATATAGCAGATTTGGAACCTTGTCTGAGACTGACTTT 2517  
 QY 2475 atcaataatcaacgagcaagaatcatatcatgltgltgltgtagacatttcaatc 2534  
 Db 2518 ATCAATAATCAACGACAAAGATCATATCATGTATGTGTAGACATTTTATTTTC 2577  
 QY 2535 atgactaacccagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2594  
 Db 2578 ATGACTAACCCAGCAGAGATTTCAGTATGCAAAATGTGTGCCCTGTGTGAGCTGAAA 2637

QY 2595 cagtcctgagccttcaaaaacctgaataagtcctccacagctgtgataaattggacaatt 2654  
 Db 2638 CAGTCTGGACTTTTCAAAAACCTTGAATGATGCTCCACAGCTGTATTAATTAATGACAAAT 2697  
 QY 2655 taggaattttaaacttagatgatacttgatgcttccatttcaatttcaatttcttctgt 2714  
 Db 2698 TAGGAATTTTAACTTTAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2757  
 QY 2715 taatgcaacagagcacttaataatgacttgatctctgttcttaagaattataaanaacat 2774  
 Db 2758 TAATGCAACAGGACTTAATTAATGAACTTGAATCTGTGTTTAAAGATTTATTAANAACAT 2817  
 QY 2775 gtcatactatacatatgctctgaggaacttagcttcaactacatacagataatgact 2834  
 Db 2818 GTGATCTATACATATGAGCTCTTGAGGACTTACCTTCACTACATCAAGATATGATCT 2877  
 QY 2835 ccatagtatccataataaactgacagagtgatltccagagtgctcgaatgttaattac 2894  
 Db 2878 CCATGATGCTCATATTAACCTGCAAGATGATTTTCCAGATGCTCGATCTGTTAATTTAC 2937  
 QY 2895 atctccatlaggctgaaagaatgactacglttctgtaatacagctgtgtcttga 2954  
 Db 2938 ATCTCCATTAGGCTGAAAGAAAGATGACCTTACGTTTCTGTATACAGCTGTGCTTTTGA 2997  
 QY 2955 tgtgtgtactgltacacagaagltgtgacgtgaggtctgctgtgtgtcgtatgaa 3014  
 Db 2998 TGTGTGTACTGTACACAGAAGTGTGTGACAGAGCTGCTGCTGTGTGTGTGTGTGTGT 3057  
 QY 3015 aaactgtgagcctgagatgaatgactgtcttccatcatgtgttgaagcgtgaaatttc 3074  
 Db 3058 AGCTGTGATCCCTGCGAGATGAATGATGCTTCAATTCATTTGATGCTGGAATTTTTC 3117  
 QY 3075 tcccatgagatgaaactgaactgaatgttctgtaataaagatgaactaaanaa 3134  
 Db 3118 TCCCATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3177  
 QY 3135 aaaaaaaa 3142  
 Db 3178 AAAAAAAA 3185

RESULT 4  
 AF061941  
 LOCUS 344 bp mRNA PRI 04-MAR-1999  
 DEFINITION Homo sapiens staufen protein (STAU) mRNA, alternatively spliced,  
 complete cds.  
 ACCESSION AF061941  
 VERSION AF061941.1 GI:4335952  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 3349)  
 Wickham,L., Duchaine,T., Luo,M., Nabl,I.R. and DesGroselliers,L.  
 Mammalian staufen is a double-stranded RNA and tubulin binding  
 protein which localizes to the rough endoplasmic reticulum  
 Mol. Cell. Biol. 19 (3), 2220-2230 (1999)  
 JOURNAL  
 MEDLINE 99147057  
 REFERENCE 2 (bases 1 to 3349)  
 DesGroselliers,L., Wickham,L. and Luo,M.  
 Direct Submision  
 Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900  
 Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada  
 FEATURES  
 source location/Qualifiers  
 1..3349  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="20"  
 /map="20q13.1"  
 1..3349  
 /gene="STAU"  
 495..1985  
 CDS

```

REFERENCE      1 (sites)
AUTHORS        Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
                Shihabara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE          NEO human cDNA sequencing project
JOURNAL        Unpublished (2000)
REFERENCE      2 (bases 1 to 3194)
AUTHORS        Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
                Shihabara,T., Tanaka,T. and Nakamura,Y.
TITLE          Direct Submission
JOURNAL        Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
                Sugano, Institute of Medical Science, University of Tokyo,
                Laboratory of Genome Structure Analysis, Human Genome Center;
                Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
                (E-mail:cdna@elms.u-tokyo.ac.jp, Tel:81-3-5449-5286,
                Fax:81-3-5449-5416)
COMMENT        NEO human cDNA sequencing project supported by Ministry of
                International Trade and Industry of Japan; CDNA full insert
                sequencing: Research Association for Biotechnology; CDNA library
                construction, 5'- & 3'-end one pass sequencing: Department of
                Virology and Human Genome Center, Institute of Medical Science,
                University of Tokyo (partly supported by Science and Technology
                Agency).
FEATURES
source          1. 3194
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /cell_line="HepG2"
                /cell_type="hepatoma"
                /clone="HEP02379"
                /clone_lib="HEP"
                /note="cloning vector pME18SFL3"
misc_feature    1. 3194
                /note="highly similar to AF061938 Homo sapiens staufer
                protein (STAU) mRNA"
BASE COUNT     894 a 787 c 732 g 781 t
ORIGIN
Query Match    93.7%; Score 2945.2; DB 34; Length 3194;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 3094; Conservative 0; Mismatches 13; Indels 81; Gaps 6;

```

```

Db 420 CTTATCCCCGAGGACTTTTACCCATTTCAGTTCCACCTTTACTTTATCAAGTGAAC 479
Qy 436 ttctgtgaggagacgaacttaatgcaaaagaagaacagagctgcgaacag 495
Db 480 TTTCTGTGGAGACAGCAATTAATGCGAAAGAACAGACAGAGCTCGCAAAACAGC 539
Qy 496 atgctgtgcgaagaagcgtttgagatcctgcgaagaatgagccctgcgaagagctggagg 555
Db 540 ATGCTGTGTCGAACCGTTGAGGATCTGCAAAATGAGCCCTTGCCAGAGAGCTGGAGG 599
Qy 556 taaatgaaagaataatcgaagaagaataatcacaataatcctaataatgaatgaatgctg 615
Db 600 TGAATGGAAGAAATCCGAGAGAAATCTCAATTAATCTAATTAAGTCAAGTGTGG 659
Qy 616 agattgcaactaaacgaagcctgtgctgtgaatttcgaagtgtgcccgaagagtgccac 675
Db 660 ACATTGCACCTTAAGGAACTTCGCTGTGAATTTGAGAGTGCCCGGAGAGTGCACAC 719
Qy 676 cccacataagaagactttgtgaccaaggtttggtgttgaggagtttgaggaggagtgagg 735
Db 720 CCCACATGAAAGACTTTGTGACCAAGGTTTCGTTGGGAGTTTGGGGAGAGTGAAG 779
Qy 736 ggaagaagaagaatlttcaagaagaatltccgcacatagctgttcttgagagctgtaaga 795
Db 780 GGAAAGGCAAGAAAGATTTCAAAGAAATGCCCATAGCTGTTCTTGAGAGCTGAAGA 839
Qy 796 agttccgcctcctgtcgtcagttgacagagtaagaacttaaaaaaagaacaaac 855
Db 840 ACTTACCGCCCTGCTGCTGAGTTGAACGATGAACCTTAAGATCAAAAGAAACAAAA- 898
Qy 856 ccatatgcaagcagacagaagaacccagaataatgtgcgaagagatcaatccgatlaacac 915
Db 899 CCATAGTCAGGCACAGACAGACCCAGAAATATGCGAGGGATCAATTCGATTAGCCGAC 958
Qy 916 tggccagatccacagcagcgaagaagaagaagcgaagcgaatgcacacagagc 975
Db 959 TGGCCAGATCCAGCAGCAGCAAAAAGAGAGAGCAGAGTACACGCTCCACAGAC 1018
Qy 976 gaagcctccgcgcgcgcgaagagtttgatgacaggttgaggtttgaaaacacactgcag 1035
Db 1019 GAGGCTCCCGCGCGCAGAGGAGTTGTGATCAGATGAAGTTGGAACCACTGCGAG 1078
Qy 1036 aaggaagcggcgcacaaagaagaagtgtgccaagcgaatgcgaacgaatgcgtgga 1095
Db 1079 AAGGAACGGGCAACAAGAAAGGTGGCCACGCAATGACCCAGAAATGCTGGAGA 1138
Qy 1096 tccttgattcaaatccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1155
Db 1139 TCGTTGATTCAAAAGTCCCGCAGCGCAGCCACCAAAACCCGACTCAAGTCAGAGAGA 1198
Qy 1156 agaacaccataaagaacacagggatggaagaagaatgaactttttgaacctggtctg 1215
Db 1199 AGACACCCATTAAGAAACCA-6GGATGGAAGAAAGTAACCTTTTGAACCTGCTCTG 1257
Qy 1216 ggaatgaataatggaatgaaataagaagatgagtttcagatgaccttaacatc 1275
Db 1258 GGGATGAATAATGGAGCTAGTAATTAAGAGATGAGTTGAGATGCTTATCTTAATGATC 1317
Qy 1276 agcagctgcctgtggaattcttcccatgtgtgccgaagtcgccagagctgtgaagat 1335
Db 1318 AGCAGCTGCTGCTGGAATTTCTCCATGCTGCCGAGAGTGCSCCAGGCTGAGAGATTA 1377
Qy 1336 gtcaagagatatacaacaaagaatttaccagagcagctccgaaatctctgcaagagcag 1395
Db 1378 GTCAAGGACATCAACCAAAAGATTTTTACCGAGGCGAGCTCGAATCTTGCAAGGCGCAG 1437
Qy 1396 taactgcacatgaacccgagaggttgatgacagggcaccctgcgcccacagcgaagca 1455
Db 1438 TAACTGCCATGATACCCGAGAGTGTGTATGGGGGACACCTCGCCACAGCCGAGACCA 1497
Qy 1456 tttaagaataacatcttcttaagccacgtaccccatgagccttcacagagacacctgt 1515
Db 1498 TTTTAAAGAAATACACCTCTTACAGGCACAGTACCCATGAGACTCTCTCAGAGACCTCTG 1557

```



QY	1186	gaaaagfbaaccttlttlttgaacctgctctbgygga tyaaaaaygactaglaataaagg	1245
Db	1261	GAAGAAGTAACCTTTTGTGAACCTGGCTGGGATGATAAAATGGGACTACTAATAAAGAG	1320
QY	1246	atgaattcaagatgacctatacttaagtaatcaagcctgctctctgaattcttcccaatg	1305
Db	1321	ATGAATTCAAGATTCCTTATCTATGAATCATGACAGCTGCTGCTGAAATTCCTTCCCATGG	1380
QY	1306	tgcccgaggtgcgcccaagctctgtagagttagtaagagaaatcaacaccaaaga ttttaacca	1365
Db	1381	TGCCCGAGGTGCGCCAGAGCTGTGAGAGATTAGTCAAGACATCAACACAAAGATTTTACCA	1440
QY	1366	gggcaagctccgaatctccgccaagggccaaggtaacctgcatagtatgcccgaagttgtgtc	1425
Db	1441	GGGCAAGCTCCGAATCTCGTCCCAAGGCCAGGTAACTCCATGTATAGCCCAAGTGTGTGT	1500
QY	1426	atggyggaacctgcgcccaagccgagaca ttttaagaatacaactcttaaggcag	1485
Db	1501	ATGGGGGACACTCGGCCCAACAGCCGAGACCTTTAAAGATTAACATCTCTTAGGCCACAG	1560
QY	1486	taaccacatggaacctctcaacgagacccctctgagcaactgtaactatcttccagatgcag	1545
Db	1561	TACCCCAATGAGACTCTCAGCAGAGACCTCTGAGCAACTGGAGTATCTTTCAGAGTCCAGG	1620
QY	1546	gattccaggtttagtaataaagaacttccccaaaaacaagaagaatltgtatctctta	1605
Db	1621	GATTCCAGGTGAATTAACAAAGACTCTCCCAAAAACAAGAACGATTTGTATCTCTTA	1680
QY	1606	tcaattgtctcttcagacccacctctgatacgaatgtatctgcaagagatgtagagctct	1685
Db	1681	TCAATTGCTCTCTTCAGCCACCTCTGTACGCCATGCTGTATCCGCAAGATGTGGAGTCTT	1740
QY	1666	gccaatgataatgctctgcgtgaaacacttaagltgctgtctgattgtggaaccaagaatla	1725
Db	1741	GCCATGATATGGCTGCGCCTGACACATCTTAAAGTGTGCTGTGAGTTGACCAACAAGTA	1800
QY	1726	cagagatagccaagaacaggaacoggaacaaatgctctgtgtgtggaggtgcttgaaccttt	1785
Db	1801	CAGAGATGCCAAGAACAGGAAGGAGCAATGTCTTGTGTGGAGGTCTCTAACCTTTT	1860
QY	1786	ctgagcatgaaacca ttaataatccaca tatactgtaaaactactgaaactgtcttga	1845
Db	1861	CTGGCATGAACCTTTTAAAAATCCACACTATATACTGAAAAATACTGAAACTGCTTTGA	1920
QY	1846	aaatttgaattctctga taacctcagtybgyccgagagaaacggtggttaaagaatgtgg	1905
Db	1921	AAATTTGGAATTTCTGTACTCTCCAGTGGGCGCAGAGACACGGTGGTAAAGAGTGGG	1980
QY	1906	cagcaagcagggaaagaacaacgaagaacacaaagagcggctctgtgycgcgtcgaactgtctg	1965
Db	1981	CAGCAGCAGGGGAACAACAAGAAACAAAGAGAGCGGCTGTGGCGGCTGGACTGTGCTG	2040
QY	1966	ggatttgttgaatgagccactcgtgtaacctcggagcgtccctagcaatagcaactcgtctt	2025
Db	2041	GGGTTTGTGTGATGGCCACTCGGTGACCTGGCGGTCTCTTAGGCATTAGCACTGCTGT	2100
QY	2026	ggggaagaagggctgtgcccaagcgaactgtgtctcccgagacacacgaacagatccacacctg	2085
Db	2101	GGGGAAGAAGGCGTGGCCACAGCTGTCTCCGGGACACAGAGATCCACACCCGTG	2160
QY	2086	ggcaactcgtgttgtgtcttlttlttccccctgtgtgaagaagaacgacgaacccct	2145
Db	2161	GGCACTCCGCTGTGTGTCTTTTTCCTTCCTGTGAAAGAAAGAGCGCACGACCCCT	2220
QY	2146	tcttaagctgtgttaactcaagaacatlttggaacaaacccctggacagcaatgtccagagag	2205
Db	2221	TCTCAAGCTGGCTACGTCAACAACATTGGAGCAAAACCTGTGACACCACTATGCCAGAGAG	2280
QY	2206	ggctttagccggcccccgagctlaaagaacacgaagaanaa tcaaatgtcttccactcaag	2265
Db	2281	GCCTTTACCGGCGCCAGAGACTTAAAGGCACAGAGAAAATCAAAATGCTTCTACTACAGC	2340
QY	2266	tgaccacaacttltctag tgtgccaagccccacacactcctctgag taaccaacacataac	2325

Db	2341	TCACCCAACTTTTTCAGTGTGGCCAGCGGCCCCACACACTCTCTGAGTACCCACACACATCAC	2400
OY	2326	caactgcttcctccccaagatgatctgtatcttgattcattatcttccttgattga	2385
Db	2401	CACGTGTTCTCTCCACAGAGTACTGTATCTTGATTGTTTCATTATTTCTTTTGATTGA	2460
OY	2386	tatgacactataaaaaatttccatttgagaattctccaatgtatcagttaaaatagcac	2445
Db	2461	TATGACACATATATAAAATTTTCATTGTGAATAATTTCCAAATTTGTATCTTAATATAGCAC	2520
OY	2446	agtttggaaacttgcctgagactgactttatcaataatcaacggcaaaagatcatatcc	2505
Db	2521	ACTTTTGGAAACTTGTCTGAGACTGACTTATATCAATATATCAACCGCAAAAGATCATATCC	2580
OY	2506	acgttatatgtgtttagacaatttatttcattpaaccaacccaagaagtttcagttgac	2565
Db	2581	ANGTGTATGTGGTTAGACATTTTATTTCATTGTAGCTAACCCAGACAGTTTCAGTGATGC	2640
OY	2566	aaattgtgtgcctcttglttcagctgtaaacagtcctgagacttcaaaaaccttgaaatg	2625
Db	2641	AAATTTGTGTGCCCTCTGTCTGAGCTGAACACAGCTCGTGACTTTCAAAAACCTTGAATAAG	2700
OY	2626	tctcccaacagtttgtataaatttggacaatttaggaattttaaactttagatgacatcttg	2685
Db	2701	TCTCCACAGTTGTATATAAATTTGACATAATTTAGGAATTTTAACTTTTGAAGATCATTTGG	2760
OY	2686	tcccatatttattcatcttcttatttctgtttaaagcaaacagagacttaaatgaacttgc	2745
Db	2761	TTCGATTTTATTTCATTTTATTTTTGTATTAATGCAACAGACGACTTAATGAACCTTTGAT	2820
OY	2746	cctcgttlttaagaatataaaaaaacatttgtatctatatacatagagctcttgagaact	2805
Db	2821	CTCTGTTTAAAGATATTTAAAAAACATTTGTATCTTATACATATGCGCTCTTGAGGACTT	2880
OY	2806	agcttccactaacactaacagatatagctcccaagtatgccaataaaccgcgagaatgc	2865
Db	2881	ACCTTCTACTACACTACAGAGATATGATGTCACATGATGTCATATTAACCTCGCAGAGTAT	2940
OY	2866	tctccagagtgctcgaatactgttaatttaacatctccatctlaagagcttgaaaaaatgcctac	2925
Db	2941	TTTCAGAGATGCGTCATACTGTTTAATTACATCTCCATTTAGGGCTGAAAAAGAGACTTAC	3000
OY	2926	gttctctatbacagctggtgtgctctttagatgtgtgttactgtatcacagaagtgtgtga	2985
Db	3001	GTTTCTGTATACAGCTGCTGTCTTGTGTGATGTGTACTGTACACAGAAAGTGTCGA	3060
OY	2986	cctgaagctcttgagtggtgtgtgcgtatagaaaaacctgttagccctcgagttgaatgcgtc	3045
Db	3061	CTGAGGCTCTGTGGTGTGGTCCGATGTGAANAACCTGTATGCGCTGCGAGTTTAAGTACTGCT	3120
OY	3046	tccattcatctgttcaacgcctgaaattcttcctcccatcgtgaagtgaatgaataaacttaagtg	3105
Db	3121	TCCATTTCATGTTTACGCGTGGAAATTTTCTCCCATGAGAAATGTAAGTAAACTTAAGTGT	3180
OY	3106	tctgcatacaataatggtataacttaaaaaaataaaa 3142	
Db	3181	TTCATCATCAATAATGTATACTATAAAAAAATAAAA 3217	
RESULT	3		
LOCUS	AK025519	3194 bp mRNA	PRI
DEFINITION	Homo sapiens cDNA: FLJ21866 f1s, clone HEP02379, highly similar to		
ACCESSION	AK025519		
VERSION	AK025519.1	GI:10438059	
KEYWORDS	oligo capping; f1s (full insert sequence).		
SOURCE	Homo sapiens heptadoma cell_line:Hepg2 cDNA to mRNA, clone_11b:HEP		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		





|||||  
Db 901 ATCCGATTACCCAGCTGGCCACATCCAGCAGCAAAAAAGAGAGAGCCAGAGTACA 960  
QY 961 cgtctctcaagaagcagagcctcccgccgcgagggaggtttgtaagcagtgtaagttg 1020  
Db 961 CGGTCTCCACAGAGCGAGGCTCCCGCGCGAGGGAGTTGTATGATCAAGTCAAGGTTG 1020  
QY 1021 gaacaacacatgacagaagagagcgcacacaagaagtgtgcgaagcgaatgcagccg 1080  
Db 1021 GAACACCACACTGCAGAGAGAGGAGGACCAACAGAAAGTGTGGCCAGAGCCAAATGCAGCCG 1080  
QY 1081 agaacaatgctgagatccttggtttcaaaagtcgcgaagcgagcccaaacccgcac 1140  
Db 1081 AGAACATGTCTGAGATCCTTGTGTTCAAAATCCCGCAGCGCAGCCACCAACCCCGCAC 1140  
QY 1141 tcaagtcagaagagagaccccaataagaaccagaggtgtgaagaagaagttaaccttt 1200  
Db 1141 TCAAGTCAGAGAGAGAGACCCCATTAAGAACCGAGGAGTGAAGAAAGTAACTTTT 1200  
QY 1201 ttgaacctgagctctgaggaatgaaatcgagactaaataagaagatgagtcaagaatgc 1260  
Db 1201 TTGAACCTGGCTCTGGGGATGAAATGGACTAGTAAAGAGAGATGATTCAGATGTC 1260  
QY 1261 cttatctaaatgcaatcagcagctgctgtgaaatctctcccatggtgcccgaagtcgccc 1320  
Db 1261 CTATATCTAACTCAATCAGCAGCTGCTGTGAAATTCCTCCATGTGTCGCCAGAGTGCACC 1320  
QY 1321 aggccttagaagtttaatcaagaagatcacacaaagatttaccagaagcgctccgaatc 1380  
Db 1321 AGGCTTAGAGAGTAAAGTCAAGAGATCACACCAAAAGTTTACCAGGGCGAGCTCCGATC 1380  
QY 1381 ctgcgaagcgaagcgaatgcaatgcaatgcaatgcaatgcaatgcaatgcaatgcaatgca 1440  
Db 1381 CTGCGCAAGCGCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1440  
QY 1441 cgaacgcgaagcgaatgcaatgcaatgcaatgcaatgcaatgcaatgcaatgcaatgca 1500  
Db 1441 CCACAGCGGAGCAGCTTTTAAAGATTAACATCTTCAGGCGCAGCTCCCATGAGACTTC 1500  
QY 1501 tcaacgaaccccttgagcaactgagactatctccagaagtcgaagatccaggttgaat 1560  
Db 1501 TCACGAGACCTCTGACCACTGACTATTTTCCAGAGTCCAGGGATTCAGAGTTGAAT 1560  
QY 1561 acaaaagactcccaaaaacaagaacgaatttgtatctctatctatctctctctc 1620  
Db 1561 ACAAGACTTCCCAAAAACAACAAGACGATTTGTATCTTTATCAATTTGCTCCCTC 1620  
QY 1621 agcgaacctgatacgaatgatacgaatgatacgaatgatacgaatgatacgaatgatacga 1680  
Db 1621 AGCCACTCTGATCAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1680  
QY 1681 cgtctgaacatcttaaaagtgtctgtctgaagtgtgaacaaagaatgacaagaatgacaaga 1740  
Db 1681 CGCTGAACATCTTAAGTTGCTGT 1740  
QY 1741 caggaacacggaacaaatgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1800  
Db 1741 CAGGAACGAGCAATGT 1800  
QY 1801 tataaaatcccaatatatactgaaatactgaaactgcttggaaaatttggaaatttct 1860  
Db 1801 TATTAATATCCCAATATATATATGAAATATCTGAAACTGTTTGAAATTTGAAATTTCT 1860  
QY 1861 gataacctcagttggccggaagacagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1920  
Db 1861 GATTAACCTCCAGTGGCGGAGAGACGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1920  
QY 1921 caacagaacacacaggaagagcgt 1980  
Db 1921 CAACAGAAACACAGAGAGGCGGT 1980  
QY 1981 gcaactcgtgtgacctgtgacctgtcctctacgcaatgacagctgtcctgtgtgtgtgtgtgt 2040  
|||||

Db 1981 GCCACTCGTGACCTGGCGGTCTCCCTACGCAATAGCAGCTGCTGTGGGGAAGAGGCTG 2040  
QY 2041 cccagccagctgtgtctccggggaacacagcagatccacacccctgtgacactcgtgtt 2100  
Db 2041 CCCAGCCAGCTGTGTCTCCGGGACACACAGATCCACACCTCTGGGACCTCTCGTGT 2100  
QY 2101 gtcctttttttccctctgtgtgaagaagaacgagcagaccccttctcaagctgtgtca 2160  
Db 2101 GGTCTTTTTCCTCGTGTGAAGAAAGAGGACGACGCCCTTCTCAAGCTGGCTCA 2160  
QY 2161 ctcaagacacatgtggacaacacccctgtgacagccatgccaagagaagccttgcagccg 2220  
Db 2161 CTCAGACACTGTGGAGCAAAACCTGTGACACCATGTCCAGAGAGAGGCTTTGACCGGCC 2220  
QY 2221 cagaagctaaagaacagagaagaaataatgctctccacacagcgtgaacacacttctc 2280  
Db 2221 CAGAGCTAAAGACACAGAGAAATCAATATGCTTCTACTACAGGTGACCCAACTTTTCT 2280  
QY 2281 agtgtgcaacgccccacacacccctctgagtaaccacacacacacacacacacacacac 2340  
Db 2281 AGTGTGACAGGCGCCACACACCTCTGCACTACCCACACCATCAGCATCTTCTCTTC 2340  
QY 2341 caacagtgatctgtatctctgaattcaatctctcttctgtatgtatgacactatata 2400  
Db 2341 CAACAGGATCTGTATCTTATGTTTCAATATTTTCTTGTATGTATGTACACTATATAA 2400  
QY 2401 aatttcaatttgaagatttctcaattgtatctatgtaataagacagtttggaaactgtc 2460  
Db 2401 AATTTTCATTTGAGATTTTCTCAATTTGTATCTAGTTAAATAGCACAGTTTGGAACTGT 2460  
QY 2461 ctgagactgacttataatcaatcaatcaacgacagaagatcaatcaatgltgtgtgtt 2520  
Db 2461 CTGAGACTGACTTATCAATATATCTAACCGACAAAGATCATATCAATGATGTATGTGTTA 2520  
QY 2521 gacatcttctatctatgactaaacccaagacagtttcaagatgacaaatgtgtgccc 2580  
Db 2521 GACATTTTATTTTCAATTTGACTTAACCGACAGCTTTCAGAGTCAAAATGTGTGCCCTC 2580  
QY 2581 tggctcagctgaacagctctctgacttcaaaaaccttgaataagcttcccaagatgta 2640  
Db 2581 TGGTTCAAGCTGAACAGCTCTGACATTTCAAAAACCTTGAAATAGTCTCCACAGTTGTA 2640  
QY 2641 taaatggaacattggaattttaaacttgaatgatactgtgttcaatcttctc 2700  
Db 2641 TAAATGGAACATTTAGGATTTTAAACTTTAGATGATCATTTGGTTCATTTTATTTTC 2700  
QY 2701 atttctatttcttgaatgaacacagacttaatgaacttgaactcgtgttttaagaat 2760  
Db 2701 ATTTTATTTTGTATATGCAACAGACTTAATGAATGAACTTTGATCTGTTTTAAAGAT 2760  
QY 2761 tatataaaacatctgtgtatctataataatgctgtgaagacttgaacttcaactacac 2820  
Db 2761 TATTAATAAACATTTGTATCTATACATATGCTCTTGAAGACTTACCTTTCACATACAT 2820  
QY 2821 acagatatgacttcaatgactatataaaccctgcagagtgatcttccagagctgtcg 2880  
Db 2821 ACAGATATGATCTCCATGATGATCAATATTAACCTGACAGATGATTTTCCAGAGTGTCTG 2880  
QY 2881 atactgttaatacatctccatagagctgtgaagaagaatgacagcttctgtatacagc 2940  
Db 2881 ATACTGTAAATATCACTCCATTAAGGCGTGAAGAAATGACCTACAGTCTTGTATACAGC 2940  
QY 2941 tgtgtgcttctgtatgt 3000  
Db 2941 TGTGTGCTTTTGTATGT 3000  
QY 3001 tggtcggtatgaaacacctgtgagccctgcagatlaagtaactgtcttccatctatgtt 3060  
Db 3001 TGGTCGATGAAACACTGTGAGCCCTGAGAGTTAAGTCACTCTTCATTCATTTGTTTGA 3060  
QY 3061 cgtctggaatttctcccaatggaatgaaacttaagtggttctgtataataataat 3120  
Db 3061 CGCTGGAATTTTCTCCCATGTGAATGTAAAGTAAAGTGTGTGTGTGTGTGTGTGTGTGTGT 3120

22	150.8	4.8	160979	52	AC055879	Homo sapi
23	149.6	4.8	33173	34	AC001576	Homo sapi
C 24	149	4.7	33281	72	ALJ30198	Homo sapi
25	133.4	4.2	194407	54	AC068845	Homo sapi
26	125.2	4.0	2008	34	AC023314	Homo sapi
27	119	3.8	2847	69	HSRNATP1	X83973 H.sapiens m.
28	110.2	3.5	5237	53	DROSADJEN	M69111 D. melanogaster
29	110.2	3.5	61267	28	AC004336	Drosophila
C 30	110.2	3.5	120625	38	AC001563	Drosophila
C 31	110.2	3.5	267271	28	AE003800	Drosophila
32	105.8	3.4	84009	29	AF225924	Drosophila
33	103	3.3	8002	29	AF225925	Musca domestica
C 34	85.4	2.7	171591	50	AC027018	Homo sapi
C 35	70.8	2.3	7218	81	166494	Sequence 14
C 36	68	2.2	7218	81	166494	Sequence 14
37	65.4	2.1	60009	48	AC023436	Homo sapi
38		2.0	180936	62	AC078936	Homo sapi
39	63	2.0	258550	35	AC010758	Mus musculus
C 40	62.6	2.0	42099	29	CELEF554	Caenorhabditis elegans
C 41	62.6	2.0	154746	80	HSV2HG52	Herpes simplex virus 2
C 42	62.6	2.0	154746	80	HSV2HG52	Herpes simplex virus 2
43	62.2	2.0	141892	47	AC023197	Homo sapi
44	62.4	2.0	89669	69	HSJ340H1	Homo sapi
45	61.4	2.0	77127	36	AC011801	Homo sapi

## IGNMENTS

[illegible]

RESULT	1
AF061940	
LOCUS	
DEFINITION	AF061940 3142 bp mRNA PRT 04-MAR-1999
ACCESSION	Homo sapiens staufen protein (STAU) mRNA, alternatively spliced,
VERSION	complete cds.
KEYWORDS	AF061940
SOURCE	AF061940.1 GI:4335950
ORGANISM	human.
	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 3142)
AUTHORS	Wickham,L., Duchaine,T., Luo,M., Nabl,I.R. and DesGroselliers,L.
TITLE	Mamalian staufen is a double-stranded RNA and tubulin binding
	protein which localizes to the rough endoplasmic reticulum
JOURNAL	Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
MEDLINE	99147057
REFERENCE	2 (bases 1 to 3142)
AUTHORS	DesGroselliers,L., Wickham,L. and Luo,M.
TITLE	Direct Submission
JOURNAL	Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900
	Edouard Montpeilf, Montreal, Quebec H3T 1J4, Canada
FEATURES	location/Qualifiers
source	1..3142
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="20"
	/map="20q13.1"
	1..3142
	/gene="STAU"
	288..1778
gene	/gene.."STAU"
	/function="RNA binding protein"
	/note="alternatively spliced transcript"
	/codon_start=1
	/product="staufen protein"
	/protein_id="AAD17532.1"
CDS	/db_xref="GI:4335951"
	/translation="MICKKMKYKVPDVSFPMOSTNYNNMGCAVPPRYAPPYPVP
	LKYVELSVGGQGFNGKTKTROAKKDAAKLRLONLPEPLERLVNGRSEENLN
	SEISVFEEALAKRLNPVNEVARSGEPFHKNVTKVSGVFGEGSKSKISKRN
	ATAIVEELKTEPLPAVERIVKPRIKKKRKIPVKPDTSPDEQGGINPSIRLAOIQAOK
	KEEPELTLLTERGLPRREFFVMQYVGHNTIEGTGNKKAKRNAAMEILTFPKV

241	ACGACGCCAAGCAAGCATACCCCTACTGTAAGACTAAATGCACTGTGCAATGGAA	300
QY	301 aaaaaccaatgtataagcctgttagacccttactctcggatgcaagtcacataactaca	360
DB	301 AAAAACCAATGTATTAAGCCTGTTGACCCCTACTCTCGATGAGTCCACATATACTACA	360
QY	361 acatagagagagtgctcttcccgcgagtaacttaccatcttcagttccactttac	420
DB	361 ACATGAGAGAGTGCTTTATCCCGAGACTGTTTACCATTTCACATTCACACTTAC	420
QY	421 ttatcaagtgaactctctgtggagagacaacatctaaygcaaaagaaagacaagac	480
DB	421 TTTATCAATGTGAACCTTCTCTGGAGGACACCAATTAATGCAAGAAAGACAACAC	480
QY	481 aggcctgcgaacaacagatgtctgtgcacaaagcttagaatacccgatagaccctgc	540
DB	481 AGGCTGCGAACAACGATGCTGCTGCCAAAGCTTTAGGATCTCGAAGATGAGCCCTGC	540
QY	541 cagagagccttgaggtgaaatggaagaatcccgaaagaanaatcctaataatctgaa	600
DB	541 CAGAGAGCTGGAGGTGAATGGAAGAAATCGAAGAAAGAAATTCATTAATCTGAA	600
QY	601 taagtaagatgttttagatgtgacttaacggaacttcccttgaaattcgagtgagcc	660
DB	601 TAAGTCAAATGTTTGAATGATTCATTAAAGGAACCTGCTGTGAATTTGAGGTGGCC	660
QY	661 gggagagtgagccaccaccaatgaagaactctgtaccaaagttctcggttggaagttg	720
DB	661 GGGAGAGTGGCCCAACCCCATGTAAGAACTTGTGACCAAGGTTTCGGTGGGAGTTTG	720
QY	721 tgggggaaggtgaagyggaagaagaagattcttaagaanaaatccgcgcataagctgtc	780
DB	721 TGGGGGAAGCTGAAGGGAAGCAAGAAAGATTTCAAAGAAATATCCGCGATAGCTGTC	780
QY	781 ttgagagcctgaagaagtataccgcctctgctgcagttgaacgagtaaaccttagaatca	840
DB	781 TTGAGAGCTGAAGAAGTATACCGCCCTCTCGAGTTGAACGAGTAAAGCTTAATCA	840
QY	841 aaaaagaanaacaaaccatagtcacagccaagaagccagataatgtgcaagggatca	900
DB	841 AAAAGAAAACAAACCCATATGTCAAAGCCACACAAAGCCCAATATATGACCGGGGATCA	900
QY	901 atccgattagccagctagtcgcagatccagtcagcaaaaagagaagaagccagagtaca	960



QY	1705	agcgcgctctgagttggcccaacagagacagagatgcgaagaacaggaatgagacc	1764
Db	1614	AGCTGCTGTCTGAGTTGGACCAACAGACAGCAAGATGCGCAAGGACAGGAACGGACAG	1673
QY	1765	tttaagcgttcgaggaggtgtctgaaccttttcgcgcacaacacattataaaccacat	1824
Db	1674	TTTCAGCGTCCGGGAGGTGCTGAACCTATTCTGGCCATGAAOCATTATAAACCCMACAT	1733
QY	1825	ataatacgaataatactgagagactgctcttgaaaaatttgaaatactgataactcaagtcg	1884
Db	1734	ATATACCGAAAAATACAGAGACTGCTTTGAAAATTGGAAATTTCGATTAACCTCAGTGGG	1793
QY	1885	ccaagaacatgtgtgat--aaaaatgtgcacaagaacgcgaagaactcagttgttagcc	1942
Db	1794	CCAAGACATGTGGATPAAAAAATGTGGCAAAAAGAAAGAAAGAACTGCAGGTGTGGCT	1853
QY	1943	tggtgtgcgtcgagcgtagtgaatgacgtgtgctctgcacatccacagacagaacacag	2002
Db	1854	TGGTGGGGCTG--GGCTGCTGGTGACGCTGTGCTCCGCCATTCa--CCAGGAGACCAGAG	1910
QY	2003	ccccaaagccctcaagttctgttttgcacgtgcgaacaa-----gagagacagccaat	2056
Db	1911	CTCCAGTGTCTCCAGCTTTTTTTTCCTCGTGACAGAGAAACGAGAGCGTGGCCAACT	1970
QY	2057	ctcactgtgctctctccagataactttgaaaaaccggacagccacacacagaagaagcccta	2116
Db	1971	CTCCAGCTGGCTCTCTTCAGATCTTTTGGAAAAACCGGGACAGCCACACAGAGAGCCCTTA	2030
QY	2117	tagcgcgcccgagcctaaagcgacagagaanaagcgacgtcttcctctacatgact	2176
Db	2031	CAGGGCCCCGGAGCTTAACGCACCCAGAGAAAA--ACAGTGTCTCT-----ACATGACT	2084
QY	2177	gactcaagcttcgcgcacagcylagca--ccaactgaacacactgtcttccttcagt-----	2228
Db	2085	GACGCAACTCTGCACACAGTAGACACCACCGTAGACACTGCTTCTCTTCAACAGTGATG	2144
QY	2229	-----ttcaatttttctctctgtattgtatacaaacacta--taatttcaattca	2274
Db	2145	GATGCCCTGTAGTTGACATTAATTCCTTTTGGTTGGAGATGACACTATGATTAATTTTCATTTCA	2204
QY	2275	gttcccttagtcggtctactactaccctagcagtttagaaatgtg-----	2316
Db	2205	GTTTCTCAGTTGGATCTACTTACTTACGTAGCATTTAGAACCTAGTCTGACACAGATGCTATC	2264
QY	2317	caagtcatgtactcgtgcgaagatatacagcccggtcttgtagt--gcaatctcgtccctcgtc	2374
Db	2265	CAGCTATGTAACTGGCAAGACACAGACGCCCTGTTGGTGATGCTCMTCTGTGCTCCTGGCT	2324
QY	2375	tgactgaacagctctggaatataccaccagaatccttgactccctgcgcctctgataaatt	2434
Db	2325	CGGCTGGCACTTGTGACATTAAACAMAAATCTTGATCTCCCAACCCCTGTGTAATAT	2384
QY	2435	ggaacgcttagagactttaaactt---agatacaaaataabgyccttttaacttat	2491
Db	2385	GGACAGCTTAGGACTTTTAACTTTAGAGATCAAAAAGATTGGTCCTTTTAACTTTAT	2444
QY	2492	ttttaaagacacacattlaaaatgaagcccttgactcttaaccataaacaag-----	2543
Db	2445	TTTTAAACACACATACCTTAAAGAGAG--CCTGACCTTTTACGATTATAACATATGTCTAAC	2503
QY	2544	-----attgtcaaaagaagtttctctgaagagta--	2575
Db	2504	GGGAGAGGTTGTACCCATTATTTAAACAGATTGGCAGTGGGAGGTCACTTGAGAGGATTTG	2553
QY	2576	-----gcttttttttatacaacacatacagaataatccgttaggcacagaagactacag	2628
Db	2564	CTCAAGGTGCTTTTCTACACACTACAGACACTTAACCTGTGGGCCCAAGAACTAATGG	2623
QY	2629	ctgagtctcccttagagggcccaatacagatcaatctcaaacctctaaatctggggaagaatgac	2688
Db	2624	CTGCTGTCCCTGGAGGGCCCAATACAGTCAATTCACACCTCCACGTA--GGAGAAGGGCAC	2682
QY	2689	aggttccctcgtgcgtg-----gtgtgcacagaagggcgaagctcagctgtgccttgggaag	2743

	Db	2683	AGGTTTCTGGTGCTGTGTCCTTAGTGTGCACANAGCAGGAAGTGTCCCGGCTGTGG	2742
Oy	2744	agcattcgctcccaatgacgcagc---ctgtcttcacactcttgattagtgaacctcc	2800	
Db	2743	AGCATTTGGGCTCGGAGTGCACACCOTGTGTGCCACATCGGGGTTTACTGTG-ACCTTTT	2801	
Oy	2801	cactcatggaaataagaataaacactcttcttgtcaccaataaagtgltaactaa	2857	
Db	2802	CACCATGGAATTATTAAGAACAATAA-GTCTTTGTCAACCAATAATGAAVACTAA	2856	
RESULT	3			
LOCUS	AF227200	3059 bp	mRNA	ROD
DEFINITION	Rattus norvegicus staufen isoform Stau+I6 (Stau) mRNA, complete cds, alternatively spliced.			
ACCESSION	AF227200			
VERSION	AF227200.1	GI:9754865		
KEYWORDS				
SOURCE	Norway rat.			
ORGANISM	Rattus norvegicus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
AUTHORS	Monshausen,M., Putz,U., Rehbein,M., Schweizer,M., Desgroesellers,L., Kuhl,D., Richter,D. and Kindler,S.			
TITLE	Two rat brain stauften isoforms differentially bind RNA Unpublished			
JOURNAL	2 (bases 1 to 3059)			
REFERENCES	Monshausen,M. and Kindler,S.			
AUTHORS	Direct Submission			
TITLE	Submitted (20-JAN-2000) University of Hamburg, Institute for Cell Biochemistry and Clinical Neurobiology, Martinistrasse 52, Hamburg			
JOURNAL	20246, Germany			
FEATURES	Location/Qualifiers			
source	1..3059			
	/organism="Rattus norvegicus"			
	/strain="Sprague-Dawley"			
	/db_xref="taxon:10116"			
	/tissue-type="brain"			
gene	/dev-stage="adult"			
	1..3059			
	/gene="Stau"			
	227..1714			
CDS	/note="RNA-binding protein, similar to Drosophila stauften; contains four RNA-binding domains: rStau+I6"			
	/codon_start=1			
	/product="stauften isoform Stau+I6"			
	/protein_id="AAP98119.1"			
	/db_xref="GI:9754866"			
	/translation="MKRPVDPHSMOSTYSTCYMGRCAYPRPTRYPPVPPLLYQVELSVGGQGFNGKMGRRPVNHDATARLRLTLOSEPRLERLVGRESSEENLNKSETSOYFETALKRNLPNEFSFPLTOVARRESGPMMNFYRVSAFGVEGEGSKISKKNAA RAVEYLORLEPRLPAVERVRPKIKKSOPKCLQTADYOCGMPI SRLAIQOAKKE KEPEYLTGERGLPRLREFFVMOVYVGHNTAEAGTNKKVAKRNAEMLILGFKVO AQRPAKLSEERTPYVKPGDGKRVYFEESPENGETNSKDEEFMRDYSHOOLDA G ILNPVPEVAQAVGYSGCHHTKDRRAAPNPAKATVTAMLRBELLGSTSPALTLS NISSGVHPGRTRPSQLTYLLSAQSFQVEYKDQFNNNNECVSLTNCSSQPLVLH GIGGDVESCHDMAALNLTKLSELDDOSTEMPRGNQPVASGCRG"			
	3043..3048			
	/gene="Stau"			
BASE COUNT	817 a 778 c 749 g 713 t 2 others			
ORIGIN				
Query Match	69.5%; Score 1985.4; DB 74; Length 3059;			
Best Local Similarity	86.8%; Pred. No. 0;			
Matches 2513; Conservative	0; Mismatches 233; Indels 149; Gaps 24;			
Oy	91	gcgcaccctctctctctagccccgcgcctgagcgcgcgccgccttcgcctccgcacatcg	150	

D	b		1	GCGCTCCCCCTTCTTGAGACCCCGGCGCGCGCCTTTCCGACCACTCCG	60
O	y	151	cctctccctcctctctggtcgagcccccttttctctgcgcgtctcaacttgcttcaccc	210	
D	b	61	TCTCTTCCTCCCTCCCGGT----CTTTCTTTTCGCGCCCTTTCTTTGGTTTT--CTTC	113	
O	y	211	ctcgcgcgccccaagaacgcgcgcgcgcggagcagctcttgggaagcagccagaagt	270	
D	b	114	CTCCGCCCGCCCGCATATACCCTCCGCGCCCGGAGACACTTGCGGGAAGCAGCCAGMAATA	173	
O	y	271	tagcttctacatattgaactaaatcatgatcttgttgaaccttgtaagaataaacatgcta	330	
D	b	174	TAGCTTTACCATTTGACCTAATGCACGTGCTGTGAATCTGAAAGAACAACAAATGTA	233	
O	y	331	aggccggtgagaccctcaactctcgcgatgcagtlccaccctacagctatgcatlgctgaagt	390	
D	b	234	AGCGCGTAGACCCCACACTCTCGATGCAGTCCACCTACAGCATATGGCATGCTGAGGTG	293	
O	y	391	cctatcccccaagatlacttltaccatcttcagtlccacccttactctlaacaagltgac	450	
D	b	294	CCATTCCCGCCCGAGATACTTTTACCCATTTCAGATTCCACGTTACTCTACCAAGTTGAGC	353	
O	y	451	tctctcgtgggcgagcacgcagtttaaagggaagaagaagtgtgagaccoccttgaacgcg	510	
D	b	354	TCTCCGGGAGGAGCACACACTTTTAATGGGAAGGAAGAAATGAGACCTGTGAACATG	413	
O	y	511	atgcgccctgcgcgttgcgcttagagctctgcagatgaacccctgcagaagaagtgttggag	570	
D	b	414	ATGCGCACTGCCCGAGCCCTTAGACACTCTGCAGAGTAGCCCTTACCAAGAGAGCTAGAGG	473	
O	y	571	taaatggaaagaagcagagagaaanaaacctcaataaalcgaaaataagccaagtlgtg	630	
D	b	474	TAAATGAGAGAGATCAGAGAGAAAGAAACCTTAATTAATCAGAAATPAGCCAAGTGTGG	533	
O	y	631	aaattgcgctggaagcggaaatttgccctgtgaatttg-----agttgg	672	
D	b	534	AGATTGGCGCTGAAGCGGAATTTTCCGTCAATTTTGACCTTTCCCTGTGACACAGGTGG	593	
O	y	673	ccggggagagtgcccaccaaccacataagaagaacttggaccaagggttcaagtltggagat	732	
D	b	594	CCCCGGAGAGTGGCCGCCCAACATATAAGAACTTGTGACAGGGTTCATAGTTGGGAGT	653	
O	y	733	ctgtagggagaggaagagggaaaaagaagaagatctccaaagaagatlgcgccagagctg	792	
D	b	654	TTGTAGGGAGAGGAGGAAGGGAAAAAGCAGAGATCTCCAAAGAAATPGCGCAAGGCGG	713	
O	y	793	tctcggagcagcttaggaagctgcacacccctccctgctgttggagcgaagtgaagccaga	852	
D	b	714	TGCTGAGGACGATMAAGAGAGCTCCACCTCTACCTGCTGTGAGCGAGTGAAGCCCAA	773	
O	y	853	tcaagaagaanaatlcagcccacactga-----aagaacgcccgaattalgtgccaagga	906	
D	b	774	TCAAGAAAGAAAAGTCAAGCCACTCTGCAGACTACAGACTGCCCCCACTATGGCAAGGA	833	
O	y	907	tgaatccatatlagtagactlgtcacagatccagccaggtcaaaaaagaagaagagccagat	966	
D	b	834	TGAATCTTATTAGAGACACTGCGACAGATCCAGAGGCCAAAAAAGAGAGAGCGCGAGT	893	
O	y	967	acatgtcctcttacagaagcagagtccttccacgtccgaaggaglttgtatgcagltaaag	1026	
D	b	894	ACAATGCTCTTAAAGAAACGAGGTCTTCCAGCTGCGCAGGAGATTTGTATGACAGGTGAAG	953	
O	y	1027	tttggagcatacacctgcagaagaggtgtgtatcaataaagaagtlggccaaggtatgtgt	1086	
D	b	954	TTTGGGCACTACACCGCGAAGGAGGAGGATACCAATTAAGAAAGGTGTGCCAAAGCGCAATGCTG	1013	
O	y	1087	ctggaacatgtctvgagatctccctgggttcaaagtctcccaagcgcagcgtcccaagcag	1146	
D	b	1014	CTGAGAACATGCTGAGAGATCTCTGGGTTCAAAAGTCCCGCAGCGCAGCTGCAAGCCAG	1073	
O	y	1147	caactcaaatcagaagaagaagctccagtaaagaacccagagacggaagaagtaagct	1206	

Db	1074	CACTCAAATTCGAAGAGAGACACTCCATAAAGAAACCGAGACGTGAGACGAAAGTAACAT	1133
Qy	1207	tttttgaaccctagccctcgggagatganaa tgnaa ctaglaacaagga cgaaggtctcaga	1266
Db	1134	TTTTTGAACCTAGGCCCTGGGGATGAATAATGGAACCAAGTATATAAGAGATGATGATTTCAGGA	1193
Qy	1267	tgccctactcttaagcaccacgaagctgtgcacgtc tgnaa tttcccaatgtgtgccggaagtgt	1320
Db	1194	TGCCCTATCTTAAAGCCATTCACAGCTGTGCACCCGGGATTTCTCCCAATGGTGGCCAGAGGTTG	1255
Qy	1337	cccaagctgtctggggttaaglaagaagacaaccaacaagaatltcaacaggagcagttccaa	1386
Db	1254	CCCAAGCTGTGGGGTTATGTCMAAGACACCATATACAAAGATTTTCGCGAGGACGCTCCAA	1313
Qy	1387	atccctgcacaaggaacggtlaactgtccatga tagcccgagagtgtgtgtla cggggacact	1446
Db	1314	ATCCGTCCAAGGCCAGCGTAACCTGCATGTATACCCGAGATTTGTTGTACGGGGCACCT	1373
Qy	1447	cgcccaacagccgagacccaatttlaagaagtaca tctcttcgaagcca cgttaaccaatgag	1506
Db	1374	CGCCCAACGCCGAGACCATTTTAAAGAGTAACATCTCTTCAGAGCACTGTACCCCATATGAC	1433
Qy	1507	cttcgcacataagacccctctgagcaactgttaactaactttccagaagcccaaggatattcaaggtgt	1566
Db	1434	CTTCGCACAGACCCCTCTGAGCAACTGTACTACTTTCCAGAGCCAGGAGATTTCCAGGTTG	1493
Qy	1567	aatacaagaatlttccccaagaacaacaagaacgagtg tlatctcactcaactgtcctct	1620
Db	1494	AATACAAAGACTTCCCAAGAACACAAAGAACAGAGTGTATCTCTCATCAACTCTCTTT	1553
Qy	1627	cacagccgctctcgttcagtcagtcagtcacgcgaagga tglvtgaagctcgtca tgaatg	1686
Db	1554	CACAGCCACTCTGTGCATGAGCATGTGCATCGCAAGAGATGTGGAGTCTCTGTCAATGATATGG	1613
Qy	1687	ctgcaacttgaagaattttaaagctgtgtctgtgagtgtgaa ccaagaagcaagaatgagcaaa	1746
Db	1614	CTGCATCTGAACATCTTAAAGCTGTGTGTGATGTGATGGACCAACAGACACACAGAGATGCCAA	1673
Qy	1747	gaacaggaanaatggaacgaatttcaagcgtgagggaggtgcgcgaacctttctgcgaacaac	1806
Db	1674	GGACAGGAAGAGGACGACGATTTCACGCTGCGGAGGTGCTGMACTATTCTGGCCATGAAAC	1733
Qy	1807	cattataaacccaactatataacttgaanaa tacttgaacactgcttbgaaatlttgnaa ta	1866
Db	1734	CATTATTAACACCCAACTATATATACCCGAAATACTGAGAACTGTTGAATAATTTGCAATT	1793
Qy	1867	tctgataactccag tgggccaagaacatgtgtgagt-- aaaaatgtgccaagaacgcgaaga	1924
Db	1794	TCTGTATTAATCTCACTGTGGCCCAAGACATGTGGATTAATAAAATGTGGCAAAAGAAAGAAAGG	1853
Qy	1925	aactcgaagtgtgagccctgtgtgtgtgtgcgtga tgaatgaatgctgtgtcctgtccatc	1984
Db	1854	AACCTGACAGTGTGTGGCTTGTGGTGGGCTG-- GGCTGCTGTGTGAACGCTGTGCTGCCCATTC	1911
Qy	1985	catccagaacagaanaaccagcccccaagcctccag tctgttlttlttgcacgtgacaacaa ----	2040
Db	1912	CA--CGAGAGACACAGACGCTCCAGGTGCTCTCAACCTTTTTTTGTGCTGCTGACAGAGAA	1970
Qy	2041	--gagagacagccaatttctcaatgtctgtgtgtcttctcagaatcttgaanaaacccggaacgc	2098
Db	1971	CGGAGAGGTGGCCCACTCTTCACGCTGGCTTCTTCAGATTTCTTTTGAAAAACCGGGACAGCG	2030
Qy	2099	cacacacagaagagcctctatagcggcccccgaggttaa caggaacagagaaaagcgaagtc	2158
Db	2031	CACACAGAGAGGCTTTCAGCGGGCCCCCGAGCTTAACGCAACCCAGAGAAAA--ACCAAGTGC	2089
Qy	2159	tctcctatgcacatgactgaactgaactcgcgcacacgttagca--cgaatgtaaaccagtct	2217
Db	2090	TTTCCG-----ACATGACTGAACGCAACTGTGCCACACGTAAGCAACCCGTAACCTACTCT	2144
Qy	2218	tctctctcaglt-----ttcaatttttctccttgaatgatatacaaac	2258
Db	2145	TTTCTCTTTCACAGTGAAGTATGCTGTATGATTCATATTATCTCCCTTTGGTGTGAGATGAAC	2204

QY	2259	ta-aaatttcattccagttcccttagtcgtygtcactaccctagagaatttgaanaact--	2314
Db	2205	TATGTAATTTTCAATTCAGATTCTTCAGTTGACCTTACTTACCTGACGAGTTTAAAGAACTAG	2264
QY	2315	-----gtcagtcacgttaacttgcgaagaatccacagccggltggtg--g	2356
Db	2265	TCTGAGACAGATGCTATCCAGCAAGTATGTAAGTACCTGGCAAAAGCAACAGCCGTGTTGGTGATG	2324
QY	2357	cattctgtgctctcgtctgtctgtgaacagtlcttgaattaccacccaagaatccttgatcc	2416
Db	2335	CNTTCTGTGCGCTCTGCGCTGCGTGGACGTTCTGGACTTTANCCACAGAAGATCCCTTGTGCTCC	2384
QY	2417	ctgcgccttgtaataatltgagacgctttagactttttaaactt---agatcaaaagatat	2473
Db	2385	CAACCCCTTGTTATTAATTGGACAGACTTtagacttttTAAACTTTAGAGATCAAAAGCTTTT	2444
QY	2474	gtctcttttaacttatttlttaagaagagacactttaaataagcccttgacttaccoca	2533
Db	2445	GGTCTCTTTTAACTTTATTTTAAAGCAGCATCTTAAAGTGAG--CCTGACCTTTACGAG	2503
QY	2534	ttataacga-----attgtccaagaagatg	2560
Db	2504	TTATATACAGATTGTCAACGCGAGAGGTTGTCACCACTTTTAAACAGATTGGCAGTGGAGG	2563
QY	2561	ttctctggaggaagt-----gctttttttaaacaactacaggaatattactgtta	2610
Db	2564	TTCATTGAGGAGATTTGTGCTCAAGGTGGCTTTTCTACACACTACAGACATTACTGTG	2623
QY	2611	ggcccaagaagactacagcgtgtgttccctagaaggcccaataagtlcaatlccaactct	2670
Db	2624	GGCCCAAGAAAGACTATAGGCGTGGTGTCCCGAGGCGCCCAATACATGCAATCCAGCCTCC	2683
QY	2671	aagtcggggaaaggtggaacgggttctctgtgtctg-----gttggaccaaggggagcag	2725
Db	2684	ACGTA-GGGAAGGGGACAGCGTTTCTGTGCTGTCTAGTGTGACACAGAGGACGAGAAGG	2742
QY	2726	tcaagctgtgccttggggaagaagcatlgtgtgtccttagtgcagcc---tttgcttccacttg	2782
Db	2743	TCTGCGCGGCTGTGTGGAGCATTTGTGGCTCGAGAGTCAACCTGTGTCCTCCACTGTGGG	2802
QY	2783	gttttagcttggaaaccttccactcaatggaatataagtaaacctcaacttcttctgtccaat	2842
Db	2803	GTTTGGCTGG--ACCTTTTCCACCCTATGAAATATPAAAGAAACTTT--GTCCTTTGTACCAAT	2859
QY	2843	aaatgtaataactaa	2857
Db	2860	AAATAGAAATACTAA	2874
RESULT 4			
LOCUS	AF061940		
DEFINITION	Homo sapiens staufen protein (STAU) mRNA, alternatively spliced, complete cds.		
ACCESSION	AF061940		
VERSION	AF061940.1		
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 3142)		
TITLE	Wickham,L., Duchaine,T., Luo,M., Nabi,I.R. and Desgroselliers,L. Mammalian staufen is a double-stranded RNA and tubulin binding protein which localizes to the rough endoplasmic reticulum		
JOURNAL	Mol. Cell. Biol. 19 (3), 2220-2230 (1999)		
MEDLINE	99147057		
REFERENCE	2 (bases 1 to 3142)		
AUTHORS	Desgroselliers,L., Wickham,L. and Luo,M.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900		
	Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada		

FEATURES	SOURCE	Location/Qualifiers
gene	CDS	1..3142 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="20" /map="20q13.1" 1..3142 /gene="STAU" 288..1778 /gene="STAU" /function="RNA binding protein" /note="alternatively spliced transcript" /codon_start=1 /product="stauflen protein" /protein_id="AA017532.1" /db_xref="GI:4335951" /translation="MKLGKKPMKPKVDPSRMOSTYVNMKGAVPRIVYPPEPPPL LYOELSVGGQGFNGKTRQAADKAADAALRLLONEPLPERLVNGRSEENLNK SEISOVELFALKRNLPVNFVARRESGPRPHKNPTTKVSVEEVEGEKSRIKNA AIAALELKLRLPLPVARERVKIRIKTKPIVKOTSPREGGINLSRIAIOQAK EKREPELILTEKGLPBRREFVMOKVGNHRAEGTGKKVAKRAAEINMELIGFPVP OROPTRKALASEETPIKKRGDDGKKRYTFPERGSDEMGTSNKEDPFMPILSHOLPA GLIDLPVEVAOVAGVSGHHDKDTPRAAPNPAAKTVTAMTARELITGGTPTMETLIK NNLSISGHVPGPLRPSEPQDYLRSRVGFQVEKDFPNKRNERNVSLINCSOPPLIS HGICKDVESHDMALNILILKLSLELDQOSTEMPRTGMGVSVGRC"
BASE COUNT	862 a	778 c 733 g 769 t
ORIGIN		
Query Match	Best Local Similarity	47.4%; Score 1354.6; DB 10; Length 3142;
Matches 1597; Conservative	Pred. No. 0;	Mismatches 269; Indels 21; Gaps 6;
Oy	23	actctcgcgaagcgctgcygggccgcgaagccgcgtcttcaacgcttttgagcgtgctgtcgtgc 82
Dd	1	ACTTCCTGCCTCCGGCGCGGAGCGCCTGAACGCGCTTCACGCTTTGGCGGCG--GCGTGGCG 58
Oy	83	gtctgtgtgcgcgtcccccctctcttcgtgag-----ccccgcgtgcyggcgccgcg 132
Dd	59	GTCCTCTCTCGGTGCCGCTCTCTTTGACCGGCTCCCCCGCGGCGGCGGCGGCGCGCC 118
Oy	133	tgcgcctcgcgcacctcgcctcttccctcctctctgtgtagcccttcttcctgcgcgtct 192
Dd	119	TGCTTCACAGGCGCACGCGCCTCTTCCCTCCCTCGCTCCCTCTTCTCCCTTTTTC 178
Oy	193	cactgtcttctaacctctcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 252
Dd	179	TTCTTCCTCTCCCTCTCTCGCGCCACC GCCCAAGACGCCCGCGCGGAGACAG--TCGG 237
Oy	253	gaaagcagccagaagtataagcttcttacatctaagctcaatgatcactgtygtgaactfg 312
Dd	238	AGCAGACAGCCGAAGAAGCATACCCCTACTGTAAGACTTAATGCACCTGTGCATGAAGACTTG 297
Oy	313	aaaagaacaacatglatagaagccgltgaaacctactctcogatlcatgcacacttaagct 372
Dd	298	GAATAAACCAATGTAATGAAGCCTGTGGACCCCTTACTCTCGATGACAGTCCACCTTA 357
Oy	373	atgagatcgybggaagtgccataccccccagatatcttaccatttccagtcacctt 432
Dd	358	ACAACATAGAGAGAGAGTCTTAATCCCGGAGGTACTTTTATCCCATTTCCAGTTCACCTT 417
Oy	433	tactctaccaagttagctctcgcgtgcyggcgagacagctttaattggaaagaaagatga 492
Dd	418	TACTTTATTATCAAGTGAAGTCTTCTGTGGGAGCAGACACATTTTAATGGCAAAGGAACACA 477
Oy	493	gaccacccgtgaacaagatgccctcgtccgtgctgagctgaagacttcgaagatgaaccc 552
Dd	478	GACAGAGCTGCGCAAAACACATGCTGCTGGCAAAAGCCTTAGAGATCTCTGAGATGAGCCCC 537
Oy	553	tggccgaaagtgtaggtlaaatgaaagaagagcgagaaagaanaactcataatcgg 612
Dd	538	TGCCAGAGAGGCTTGAGGTGAATGGAAGAGATCCGAAGAAGAAAATATCTCAATTAATCTG 597

```

Qy 613 aaataagcgaagtgtctgaatctgcgtcgaagcgaattgcctgtgaatttgagtg 672
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 598 AAATAAGCAAGTGTGTTGAGATTGCATTAAACGAGACTGCTGTGATTTCCAGATG 657
Qy 673 cccggagagtgccacacacacacacacacacacacacacacacacacacacacac 732
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 658 CCGCGGAGAGTGGCCACCACCACATGAAAGAACTTTGTGACCAAGCTTTGGTGGGAGT 717
Qy 733 ttgtgtaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 792
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 718 TTGTGGGGAGAGTGAAGGAGAAACAAAGAAATTTCAAGAAATATGCGCCATAGCTG 777
Qy 793 ttctgagacagcttagagagctgcaccctccctcctcctcctcctcctcctcctc 852
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778 TTCTTGAGAGAGCTGAAGAAGTTACGGCCCTGCTGCAAGTTAAGAGATGAAGCCTAGA 837
Qy 853 tcaagaagaagaagltacagccacctgca-----agacagccccgattatgccaagga 906
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 838 TCMAAAGCAAAACAAACCAATAGTCAAGCCACAGACAAAGCCCAATATGCGCCAGGGGA 897
Qy 907 tgaatccattatagtagacttgtagcagatccagcagaagaagaagaagaagaagaagt 966
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 898 TCAATCCAGATTAGCCGACTGGCCAGATCCAGACGCAAAAAGAGAGAGCCAGAGT 957
Qy 967 acatgctccttaacagaagagcttccacagctgcagagagtggttgatgcagtaaga 1026
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 958 ACAGCTCTCTTCACAGAGGAGGCGCTCCGCGCGCGAGGAGTGTGATGACGATGAAG 1017
Qy 1027 ttgagcatcaacatgacagagagtggtggtacacataagaagtgagcgaagtgatgctg 1086
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1018 TTGGAAACCACTGCTCAGAACGACGACACCAACAGAGAGTGGCCAAAGCCCAATGCGAG 1077
Qy 1087 ctgaagaacatgtgtgagatccctcggggttcaaaagtccccagcagagcttcgaagcag 1146
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1078 CCGAACAACATGCTGGAGATCCTGTGGTTCAAAAGTCCCGCAGCGGACCCACCAACCCG 1137
Qy 1147 cactcaaaatcagaagaagaagacttcagtaagaagaacagaagaagaagaagaagaagt 1206
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1138 CACTCAAGTCAAGAGAGAGAAAGACACCCATTAAGAAACCAAGGAGTGAAGAAGTAACCT 1197
Qy 1207 ttttgaacctagccctcgtggatgtaaaatggaactagttaacagaagcagaagatccaga 1266
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1198 TTTTGAACCTGGCTCTGGGGATGAAATGGCATGTAATTAAGAGATGAGTTCAGGA 1257
Qy 1267 tgccttactcttagcctacagcagctgcagcagctgcgaatctcccaatggtccggaagt 1326
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1258 TGCCATTACTTAAGTATCAGCAGCTGCTGCTGGAATTTCTCCATGCTGCGGAGCTCGA 1317
Qy 1327 cccaagctgtcgggttagtcaaggaacacacacacacacacacacacacacacacac 1386
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1318 CCCAGGCTGTAGAGAGTGTAGTCAAGGACATCACCAAGATTTTACAGGAGCTCCGA 1377
Qy 1387 atcctgcagaagcaacgtaactctgcatgtagcctcggagagtggttgtaaggggacact 1446
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1378 ATCTGCGCAAGGCCAGGTAACTGATGATAGCCCGAGAGTGTGTATGGGGGACCT 1437
Qy 1447 cggccacagcagaagacatttaagaagtaaacatcttccaagcagactacccaatgac 1506
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1438 CCGCCACAGCCGAGACATTTTAAAGATTAACATCTCTTCAAGCCACGATACCCCATGAC 1497
Qy 1507 ctgcactagacaccttgagaacactgtactacttccaagacccaaggaatccagatgtg 1566
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1498 CTCTCAGCAGACCTCTGAGCAACTGAGACTATCTTCCAGAGTCCAGGATTCAGAGTTG 1557
Qy 1567 aatacaagaatlttcccaagaacaagaagaagatgtgtatctctcaactcctcct 1626
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1558 AATACAAAGACTTCCCAAAAACAAAGAACGAATTTGTATCAATTCCTCT 1617
Qy 1627 caaagcctcctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1686
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1618 CTCAGCCACCTCTGTATCAGCCATGATGATGCAAGAGATGTGAGTCTCCCTATGATG 1677
Qy 1687 ctgcactgaacatlttaagctgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1746

```

```

Db 1678 CCGCCCTGAACATCTTAAAGTGTGCTGTGAGTGTGACCAACAAAGTACAGATGCCAA 1737
Qy 1747 gaacaggaatgagaccagcttcaagcgtcggagagtgctgcgaacctttctgcacacac 1806
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1738 GAACAGGAACGCAACGCAATGTGTGTGGAGGTGCTGAACCTTTCTGTGCCATGAAC 1797
Qy 1807 catlaaana-cccaacatatactgaataatcagaagacgtccttgaanaattggaat 1865
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1798 CATTATTAATTCACATATATACTGAAATATCTGA-AACTGCTTTGAAATTTGGAAAT 1856
Qy 1866 atctgataactccagtgggccaagaca 1892
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1857 TTCTGATACCTCCAGTGGCGCAGAGA 1883
RESULT 5
AF061938 3217 bp mRNA PRI 04-MAR-1999
LOCUS AF061938 Homo sapiens stauufen protein (STAU) mRNA, alternatively spliced,
DEFINITION complete cds.
ACCESSION AF061938
VERSION AF061938.1 GI:4335946
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 3217)
AUTHORS Wickham,L., Duchaine,T., Luo,M., Nabl,I.R. and Desgroselliers,L.
TITLE Mammalian stauufen is a double-stranded RNA and tubulin binding
protein which localizes to the rough endoplasmic reticulum
Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
99147057
JOURNAL
MEDLINE 2 (bases 1 to 3217)
REFERENCE Desgroselliers,L., Wickham,L. and Luo,M.
AUTHORS Direct Submission
TITLE Submitted (29-APR-1998) Blochemistrty, Univrsity of Montreal, 2900
JOURNAL Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada
FEATURES
source location/Qualifiers
1..3217
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.1"
1..3217
/gene="STAU"
363..1853
/gene="STAU"
/function="RNA binding protein"
/note="alternatively spliced transcript"
/codon_start=1
/product="stauufen protein"
/protein_id="AADI7530.1"
/db_xref="GI:4335947"
/translation="MKLQKKPMYRPVDSRMOSTYNYNMKGAVPPRYEYPPVPL
LYOVELSVGGQFNGKGTROAKHDAKALRLILONEPLRELEVNGRESSEENLNK
SEISOVEIALKRNLPVNEEVARSGPHEMPTVKYSGFVEGSEKSKIKNA
AIAVLEIKTLPPIPAVERVKPRIKKTKTPVKPOTSPYEGOGINPISRLAIOOAKK
EKPEPTILTERGILPRRREYVQYKVGHNHREEGTGNKKVAKRNAEMLETFKYP
QROPKTPALKSEKPTPIKPKGDGRVTFEEGSDENGTSSKKEDEFKMPYLSHOOLPA
GILPMVEVADAVSOGHHHTKDETRAAPNPAKATVAMIAKRELLYGTSPTATILK
NIISSGVHPGLTRPSOOLDYLSRVGFGVEYKDFPKNNNEFVSLINCSOPLIS
HGICKDVESCHDMAALNLIKLSLELDOSTMPRTGNGPMNSVCGRC"
BASE COUNT 889 a 796 c 742 g 790 t
ORIGIN
Query Match 44.4%; Score 1269.8; DB 10; Length 3217;
Best local similarity 81.1%; Pred. No. 0;
Matches 1590; Conservative 0; Mismatches 277; Indels 94; Gaps 6;
Qy 23 acttcctgccaagctgcggcccgagccgctctcagcglttcgctgcgtcgtcgtcgcac 82

```



Db	1	ACTTCTCGCGGGGCTGCGGGCGGCTGAGGCGCTTTCAGCGTTTGGCGGC--GGCTGCGC	58
OY	83	gtctgtgtgcgtcccccctctctgtag-----ccccgcgtgcygagccgcgc	132
Db	59	GTCCTCTCGGGCTCCGCGCTCTCTTTGACGCGCTCCGCCCCCGCGGCGCGCGCC	118
OY	133	ttagcctcgcacatcgcgcctctccctcctcctgtgtggcccccttltccgcgcgtc	192
Db	119	TCTCCACGGCGCACCTCGCGCTCTTCCCTCCCTTCGCTCCCTTCCTCCCTTTTTC	178
OY	193	cactgtcttctaacctccctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	252
Db	179	TTCTTCTCTTCCCTCCTCGCGCGCACCGCCAGACGCGCGGCGGAGAGTCCGGA	238
OY	253	ggaagcagcc-----	262
Db	239	GACACACCCAGAGTTTATTACACTTAACCTCAGAACTGAACAAAGACACATTGTT	298
OY	263	-----agaaatlatagcttctacatltgagctcaatgcact	298
Db	299	CGTGAACGCCCTCTTTTAAAAAAGAAAGCATTAACCCCTACTGTAGAACTAATGCACT	358
OY	299	gtgtgtgaacctgaaagaaacaaatgtatagacocgtggaacccctcctctgtgatgca	358
Db	359	GTGATGAACCTTGGAAAAAACCATTGTATAGCTGTGACCTTACTCTCGAGATCA	418
OY	359	gtccacctacagctatgcatgctgtgagtgctctaccccccaatacttlaacat	418
Db	419	GTCACCTATTAACTACAAACATGAGAGAGTGCTTATCCCGGAGAGTACTTTACCAT	478
OY	419	tccagtcaccaacttactctactaccaagtltgagctctcgttggcgcagacagatlaatg	478
Db	479	TCCAGTTTCCACCTTACTTATTAAGTGAACCTTCTGTGGGAGACAGCAATTTAATGG	538
OY	479	gaaaggaagatlagaccaccggtgaagaacatgccccctgcgtgagcgcgaagacct	538
Db	539	CAAAAGAAAGACAAGACAGGCTCGAAACACATGCTGTGCAAAAGCGTTGAGGATCT	598
OY	539	gcaagtgaaacccctgtgcacaagaagctltgaggtlaaatlgaaagagaagcagaaga	598
Db	559	GCAAGATGAGCCCTTGGCAGAGAGGCTGAGAGTGAATGGAAGATACTCGAAGAA	658
OY	559	cctcaataaactcggaataagccaagtgtltgaaatltgcgtcgtgaagcggaaattgcgt	658
Db	659	TCTCATTAATCTGAATTAAGTCAAGTGTGTGAGATTGCATTAAAGGAACCTTGCGCT	718
OY	659	gaatttgaagtgccgcggagagtgtgcccacacatlgaaatltgtgaccaggt	718
Db	719	GAATTTGAGAGTGGCCCGGAGAGGTGGCCACCACCACATGAAGAACTTGTGACCAAGT	778
OY	719	ttcagttggggaattlttaggggaaggaaggaaggaaggaaggaatctcaagaagaa	778
Db	779	TTCCGTTGGGAGTTTGTGGGGGAAGGTGAAGGAAACCAAGAAATTTCAAAGAAAA	838
OY	779	tgcgcgaagagctgttctctgagcaggtltgagagagctgtgcaacccctcctgtgtgaagc	838
Db	839	TGCGGCCATAGCTTTTGTAGGAGCTGAAGAAGTTACCGGCCCTGCTGCAGTTGAAGC	898
OY	839	agtgaaagccaagaatcaagaagaagtgcaagccacttgc-----agacagcccgga	892
Db	899	AGTAAAGCGCTGATTCAAAAAGAAAAACCAACCATATGCAAGCCACAGACAGCCAGA	958
OY	893	ttatggccaagagatgatactctatgttagacttgcacagatctcagcagggcaaaaaagga	952
Db	959	ATATGGCCAGGGGATTCATCCGATTCGATGCGCACTGGCCCGCATTCACAGCGCAAAAAAG	1018
OY	953	gaagagccaagatacatgtctcttcaagagaagagtccttcaagctgcgaagagattgt	1012
Db	1019	GAAGAGCGCAGATGACACGCTCTCTCACAGAGCGAGGCTCCCGGCCCGCAGGAGATTGT	1078
OY	1013	gatgcagataaagcttgggcatcacatctgcagaagagtggttaccaaataagaagtggc	1072

[illegible]

TITLE Mammalian staufen is a double-stranded RNA and tubulin binding protein which localizes to the rough endoplasmic reticulum  
JOURNAL Mol. Cell. Biol. 19 (3), 2220-2230 (1999)  
MEDLINE 99147057  
REFERENCE 2 (bases 1 to 3506)  
AUTHORS Desgroselliers, L., Wickham, L. and Luo, M.  
TITLE Direct Submission  
JOURNAL Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900  
Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada  
FEATURES  
Source  
1..3506  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="20"  
/map="20q13.1"  
1..3506  
/gene="STAU"  
409..2142  
/gene="STAU"  
/function="RNA binding protein"  
/note="alternatively spliced transcript"  
/codon\_start=1  
/product="staufen protein"  
/protein\_id="AADI7531.1"  
/db\_xref="GI:4335949"  
/translation="MSOVYOVONPSAALSGSOLLNNKNSLLSOPIMSIPTSSLP  
ENMRPIONSALPASTSTSAABESTIPYELNALCKLKKRMKRYDPSYMSOST  
YINRMGGATPRTYFPPEPLLYOVELSGGQFNGKTRQDAKHAADAALRL  
QNEPLPERLENGRESEENLNKSEISQVEFIALKRLPVNEVAREGPRHMFV  
KVSVEGEVEGEGSKRSKKNAAIVLEELKLPPLPAVRVAPRIKKKPKVQ  
TSPYGGGINPISRLAOIQOAKKEPEYTLTERGLPRREPMQYVGNHAEGT  
TNKVKARNAENMLLEIGFVKPOROPTRPKLSEKTRPIKPDGKRVPFEPSGD  
ENGTSNKDEPRMPLYSHOOLPAILMPVPAVAGAVCSGCHTKDTPRAAPNAKAT  
VTATIAELIXGTSPTAETILKNNISSGHPRLTPRSOLYLSRVQSFQYERK  
FPKNKMEFVSLINCSSOPLISHIGIGDVESCHDMAALNLIKLLSELDQSTEMPT  
GNMPMSVCGRC"  
BASE COUNT 965 a 875 c 790 g 876 t  
ORIGIN  
Query Match 44.0%; Score 1256; DB 10; Length 3506;  
Best Local Similarity 86.7%; Pred. No. 0;  
Matches 1420; Conservative 0; Mismatches 210; Indels 8; Gaps 3;  
QY 262 cagaagaatagctctctccatctgagctcaatcgtgtgtgtaaacctggaagaagaac 321  
DB 611 CAGAAAGCATTAACCCCTACTAGAACTAATGCACTGTGATGAACCTTGAAAAAAG 670  
QY 322 caatgtataagcccgctgaaccctcactctcgtatgacgtaccactacagctatggcatgc 381  
DB 671 CAATGTATAAGCCTTTGACCCCTTACTCTCGATGCACTCCATATACATAACATGA 730  
QY 382 gtggaggtgctactccccccagatactttaccatctccagctccacaccttacttacc 441  
DB 731 GAGGAGGCTTATCCCGGAGAGTACTTTACCACTTTCACATTCACACTTACTTATC 790  
QY 442 aagttgaactcctcgctggcgagacaggttaatggaagaagaagtgaagaccacgcg 501  
DB 791 AAGTGAAGCTTCTGTGGAGAGACAGCAATTTAATGGCAAGAAAGAAAGAGAGCGTG 850  
QY 502 tgaacacagatgccccctgcgcctgagagcctctgacagctgaacccccgcagagaa 561  
DB 851 CGAAACAGCATGCTCTGCTCCAAAGCGTTGAGAGATCTCTCAGAAATAGCCCTCGCAGAGA 910  
QY 562 ggttggaagtaaatggaagagagagaggaagaacacacctaataatcogaataaagcc 621  
DB 911 GGCTGAGAGTGAATGAGAGATCCGAAAGAAATCTCAATAATCTGAATAATAGTC 970  
QY 622 aagtgcttgaataatgctggaagcggaattgctgctgtaatttgaagtgtgcccggagaa 681  
DB 971 AAGTGTGTTGAGATTGCACTTAACGGAACCTTGCCTGTGAATTTTCAGGTGGCCGCGGAGA 1030  
QY 682 gtggcccaacacacatgagaacttctgacacagggttcagcttgagggaatttgtagagg 741

DB 1031 GTGGGCCACACCACATGAAGAACTTTGTGACCAAGTTTCGTTGGGAGTTGTGGGG 1090  
QY 742 aagagaagggagaagaagaagaatctccaagaagaatgycgcagagctgtctcgaagc 801  
DB 1091 AAGGTGAAGGGGAAAGCAAGAAAGATTTCAAAAAAATGCCCATAGCTGTGTTGAGG 1150  
QY 802 agcttaagaagctgcacccctccctcgtctgtggaagcgagtgagaccagaatccaaga 861  
DB 1151 ACCTGAAGAGTTACCGCCCTGCTGCTGCACTTGAACGAGTAAGGCTTAGAATCAAAAA 1210  
QY 862 aaagtcagcccaactgcac 915  
DB 1211 AAACAAAACCCATAGTCAAGCCACAGACAAAGCCAGAAATATGCGCAGGGATCAATCCGA 1270  
QY 916 ttagtagaacttgcacagatccagcagcagcaaaaaagagaagagccagagatcagtcc 975  
DB 1271 TTAGCCGACTGGCCCGAGATCCAGAGGCAAAAAAGAGAGAGGAGGAGAGTACAGCTCC 1330  
QY 976 ttaacagaacgaagctctccacgtcgcaggaagttgtgtagtgagtgaaagttggaatc 1035  
DB 1331 TCACAGAGCGAGGCTCCCGCCCGCAGGAGTTGTGATGCGAGTGAAGGTTGGAAACC 1390  
QY 1036 acaactgcagaagagtggtgtaccaaataagaagtgagcgaagcgtatgctgtggaaca 1095  
DB 1391 ACACGTGCAAGAAAGAACGGGACCAACAAAGAGTGGCCAAAGCGCAATGACGCCGAGACA 1450  
QY 1096 tgcgtgagatcctggtgtcgaagtctccacagcgacgctgcgaagccagctccaat 1155  
DB 1451 TGCTGAGATCCTTGTTTCAAGTCCCGCACCGGACGCCCAACACCCGCACTCAAGT 1510  
QY 1156 cagaagaagaagctcagttaagaagaacagagaagagaagaagtaacgtttcttgaa 1215  
DB 1511 CAGAGGAGAGACACCATTAAGAAACCGAGGAGTGAAGAAAGTAACCTTTTGAAC 1570  
QY 1216 ctgacccctgggagatgaataatggaactagtaacaagaagcagaggttcagatgcct 1275  
DB 1571 CTGGCTGCGGGATGAATAATGAGTGAATGAAGAGATGAGTTCAAGATGCTTATC 1630  
QY 1276 ttaagcacaagcagatctgcagctggaattctcccaatggtgcgaagtgtgcagagctg 1335  
DB 1631 TAAAGTATCAGCAGCTGCTGCTGGAAATTTCCATGCTGCGCAGGTCGCCAGGCTG 1690  
QY 1336 tccgggttagtcaagaagacacacacacacacacacacacacacacacacacacac 1395  
DB 1691 TAGGAGTTAGTGAAGACATCACACCAAGATTTTACAGGAGACTCGGATCTCGCA 1750  
QY 1396 aggcacaggttaactgcatgataagccagagatgtgtgtaacggggacactgcgccaag 1455  
DB 1751 AGGCCACGGTAACTCCATGATAGCCCGAGAGTGTGATGGGGGACACTGCGCCACAG 1810  
QY 1456 ccgagaacattttaagaagtaacatctcttcagcgaagctacccatgagctgcgacta 1515  
DB 1811 CCGAAGACATTTTAAAGATATCAATCTTTCAGGCGACCTACCCATGAGACTCTCAGCA 1870  
QY 1516 gacctctgaagaactgtactactcttccagagccagagatctccaggttgataaaga 1575  
DB 1871 GACCTCTGAGCACTGCACTATCTTTCCAGAGTCCAGGAGATTCCAGGTTGAATCAAA 1930  
QY 1576 atttcccaagaacaagaagaagtgltatctctcaacactgctctcacagcgc 1635  
DB 1931 ACTTCCCAAAAACAAAGAAAGAAATTTGTATCTTATATCAATGTGCTCTCAGCCAC 1990  
QY 1636 ctctcgtcaagtcatagtgcatacgcaagatgtgagatctctgtcataatgctgcacga 1695  
DB 1991 CTCTGATCAGCCATGTTATCGCAAGATGTGAGTCTCGCATATATGCTGCGCTGA 2050  
QY 1696 acattttaagctgctgctcgtgagtgagcacaagagacagacagagatgccaagaacagg 1755  
DB 2051 ACATCTTAAAGTGTGCTGTGAGTTGACACAAAGTACAGATGCCAAGACAGCA 2110  
QY 1756 atgagacagttcgaagcgagagagtgtagaaccttcttgccacaaacattataa 1815

QY	1816	a-cccaacatatactggaacacgtcttggaacatttggaatactgataa	1874
Db	2171	ATCCCAACATATATACCTGGAATAACTGTA-ACCTGCTTTGAAATAATTGGAAATTTTCGATAC	2229
QY	1875	ctcgaatgggccaagaca	1892
Db	2230	CTCCAGTGGGCCGAGAGA	2247
RESULT 7			
AF061941			
LOCUS	AF061941	3349 bp	mRNA
DEFINITION	Homo sapiens staufen protein (STAU) mRNA, alternatively spliced, complete cds.		PRT
ACCESSION	AF061941		04-MAR-1999
VERSION	AF061941.1	GI:4335952	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 3349)		
TITLE	Wickham, L., Duchaine, T., Luo, M., Nabi, I. R. and DesGroselliers, L. Mammalian staufen is a double-stranded RNA and tubulin binding protein which localizes to the rough endoplasmic reticulum		
JOURNAL	Mol. Cell. Biol. 19 (3), 2220-2230 (1999)		
REFERENCE	99147057		
REFERENCE	2 (bases 1 to 3349)		
AUTHORS	DesGroselliers, L., Wickham, L. and Luo, M.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-Apr-1998) Biochemistry, University of Montreal, 2900 Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada		
FEATURES			
source	Location/Qualifiers		
gene	1..3349		
CDS	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="20"		
	/map="20q13.1"		
	1..3349		
	/gene="STAU"		
	495..1985		
	/gene="STAU"		
	/function="RNA binding protein"		
	/note="alternatively spliced transcript"		
	/codon_start=1		
	/product="staufen protein"		
	/protein_id="AAD17533.1"		
	/db_xref="GI:4335953"		
	/translation="MKGKKPMYRPVDPYSMSOSTYNNMKGAYPPRYFVPPVPLI LYVELSYGGOGFNGCKGTROAKHDDAAKRLRLONPLPERLEVNGRSEEFENLNK SEISQVEPILAKRLNDPVNFVARESGPRPHMKNPTTKSVGFEVDEBKSRRISKNN AIVALEIKRLPLPVAERVKPRIKTKTPIAKVQTSPEYGOGINPISRLAQIDQARK EKBEYVLLTERGTPRRRRFVQYKVGNHTEGTYNKKVAKRNAENMELIGFKVYV OROTRPLAKSEETKPIKPGDGRKVTFFEGSDENGTSNKDEFRMPYLSHOOLPQ GLIPVPEVAOAVGSGHHKCDTPRAAPNPAKATVAMIARELLYGGTSPTAETILK NITSSGHVPHGPILRPSOLDYLSRVGFOVEYEDFPRKNKNREYSILNCSSQPLPIS HGIGKDVSCDMAALNLIKLISETDOSTEMPTRGNGPMVSGRC"		
BASE COUNT	913 a 838 c 777 g 821 t		
ORIGIN			
Query Match	43.9%	Score 1255	DB 10; Length 3349;
Best Local Similarity	86.7%	Pred. No. 0;	
Matches 1419; Conservative	0;	Mismatches 210;	Indels 8; Gaps 3;
QY	263	agaagatagcttcacattgagctcagctcagctgctgtggaaclyggaagaagacc	322
Db	455	AGGAAGCATTAACCCCTACGTAGTAACTAATGCATCTGTGCATGAACCTTGGAACCAAAAC	514
QY	323	aatgtataagcccgtygagccctcactctcgtatgcaatgcacataagctatgcatgcg	382

Db	515	AATGTATAAAGCCTGTTGACCCCTTACTCTCGGATGCAGTCCACCTTTACTTACACATGAG	574
Qy	383	tgagagtgacctatccccccagatacttaccatttccagtcaccacttacttacc	442
Db	575	AGGAGGTGCTTATTCGCCGAGGTACTTTTACCATTTCCAGTTTCACTTACTTTATCA	634
Qy	443	agttgagctctcgttggcgagacagttttaatvggaaaggaagatvgaacccgct	502
Db	635	AGTGAACCTTTCTGTGGGAGGACAGCAATTTATATGCGCAAGGAAAGACAAACAGGCTGC	694
Qy	503	gaacacagatgccccctggccggttgcgcttggagctctgacagatgaaacccctggcagaa	562
Db	695	GAACACGATCTCTGCTCCAAAGCGTTGAGATCTGCGAATATGAGCCCTTCGACAGAG	754
Qy	563	gttggagtgtaattgaaagagaagcaaggaagaaaccctcaataatcggaataaagca	622
Db	755	GCTGGAGGTGATGTGAAAGAAATCCGAAAGAAATCTCATTAATCTGAATATAGTCA	814
Qy	623	agtgcttgaatattgctgcttgaagcgaatttgccttgaattttsagttgcccggagag	682
Db	815	AGTGTGTGAGATTTCCTTTAAAGGAACTTGCTGTGATTTTCGAGGTGGCCCGGAGAG	874
Qy	683	tgggccaccaccacatgaaagactttgtgaccggagtttcaagtttgggaattttgtaggga	742
Db	875	TGGCCACCCACACATGGAAGACTTTGTGACCAAGTTTCTGTTGGTGGGAGTTTGGGGGA	934
Qy	743	aggagaagggaagaagcaagagaatctcccaagaagaaatggcgccagagcgttcttgaagca	802
Db	935	AGGTGAAGGGAAGGAAAGCAAGATTTTCAAGAAAAATGCCGCATATGCTTTCTTGAAGA	994
Qy	803	gctttagagagcttgcacacccctccctctgttgyagcgaagtgaagcccaatcaagaagaa	862
Db	995	GCTGAAGAAGTTTACCGCCCTGCTGCGTAGTGAACGATGAAGCTTAGATCAAAAAAGAA	1054
Qy	863	aagtcagcccaactcga-----agacagcccggaattatgyccaaggatgtaactctat	916
Db	1055	AACAAAAACCATTAAGTCAAGGCACAGCAACACCCAGATATTGGCCAGGGGATCAATCCGAT	1114
Qy	917	tagtgaacttccacaagatctccagcagccaataaagagaagagccagagatcactactct	976
Db	1115	TAGCCGACTGGCCCAATTCACGACGCGCAAAAAGGAAAGGAGGCCAGATACAGCTCTCT	1174
Qy	977	tacagaacgaggtcttccagctctgcagggagtttgtatgctcagtlaaagtttgygcaltca	1036
Db	1175	CACAGAGGAGGAGCCTCCCGGCGGACAGGGAGTTTGTGATGACAGTGAAGGTTGGAACCA	1234
Qy	1037	cactgacagaagagtggtgtaccaaataagaagtgtgcacaaagcgttaatgctgtctgaataat	1096
Db	1235	CACCTCCAAAGAAAGCGGGCACCAACAAAGAGGTGGCCAAAGCCCAATGACGCGAAGACAT	1294
Qy	1097	gctggaagatctctgggtgttcaaaagtctcccgagcgagacgcttgcgaagcagaactaaatc	1156
Db	1295	GCTGAGATCTCTTGTTTCAAAAGTCCCGCAGCGGACGCCACCAAAACCCGACACTAAGTC	1354
Qy	1157	agaagaagaagctctcagtlaaagaacacagagagaacggaaggaaglaacglttltbgaac	1216
Db	1355	AGAGGAGAAGACACCATTAAGAAAGAACAGGCGATGGAAGAAAGATACTTTTGAAC	1414
Qy	1217	tagccctvgggatgaaataatggaactagtaacaagaagcagggagttcagagatgcttactat	1276
Db	1415	TGGCTCTGGGGATAAATATGGGACTAGTATAAAGAGGTAGTGTTCAGATGCTCTTATCT	1474
Qy	1277	tagcctaagaagcttgcacagctgaatcttcccacatggtgtgcgcggaagtltgccacagctgt	1336
Db	1475	AAGTCATCAGCAGCTGCCCTGCTGGAATTTCTCCATGTGTGCCGAGTGTGCCACAGGCTGT	1534
Qy	1337	cggggtttagtcaagaagacacacaccaaagaatttcaaccagggcagctccaatctctgcca	1396
Db	1535	AGGAATTAGTCAAGACATCAACACAAAGATTTTACACAGGACAGCTCCGAATCTGCGAA	1594
Qy	1397	ggcaacggttaactctcattgatagcccgagagtggtgtctcaggggcaactctgcccacagc	1456
Db	1595	GGCCACGGTATCTCCATGTATAGCCGAGAGTGTGTGTGGGGGACCTGTGCCACACGC	1654



QY	812	gctgcacccctccctcgcctgctgagagatgaagcccaagatcaagaagaatgcagc	871
Db	841	gTTACCGCCCTCTCGTCACTTGAACGACTAAAGCTTGAATCAAAAAGAAACAAACC	900
QY	872	-----caactgcagaagcagcccccgaattatgcgaaggatgtaacctattagtagact	926
Db	901	ATAGTCAAGCCACAGACAAGCCAGAAATATGGCCAGGGGATCAATTCGATTAGCCGACTG	960
QY	927	gcacagatccagcagcgaacaaagaagaagacagatgacatgctccatagaacga	986
Db	961	GCCAGATCCAGACAGCCAAAAAGGAGAGGCCAGAGTACAGCTCTCCACAGACGA	1020
QY	987	ggtctccacgtccaggggaatttgatgacggtgaaggtttggcatcacacgtcagaa	1046
Db	1021	GCCCTCCCGCGCGAGGAGTTTGATGACAGGTGAAGGTTTGGAAACACACCTGCAGAA	1080
QY	1047	ggaatggtaccataaagaagtgtgcacagcgtaaigtctcgtgaacatgctgaatc	1106
Db	1081	GGAACGGGCACACAAGAAAGTGGCCAAAGCCCAATGCAGCCGAAACATCTTGAGATC	1140
QY	1107	ctgggtgtcaagtlcccccagcgcacgctgcgaagccagcagcactcaatcagaagaag	1166
Db	1141	CTTGTTTCAAAATCCCGCAGGCGAGCCACCAAAACCCGCACTCAATGACAGAGAGAG	1200
QY	1167	actccagtaagaagaccagagagacggaagaagtacagtttttgaacctagccctggg	1226
Db	1201	ACACCCCTAAAGAAACAGGAG-GATGGAAACAAAAGTAACTTTTGAACCTGGCTGGG	1259
QY	1227	gataaaatggaactagtacaagaagcagaggttcagaatgacctatcttagcatcag	1286
Db	1260	GATGAAAAATGGACTAGTAAATTAAGAGAGATGAGTTTCAGATGCCATTACTTAAGTCAATCAG	1319
QY	1287	caagctgcagctgynaaltctccccaatgltgcgggaagttgcaccaagctgctggggttagt	1346
Db	1320	CAGTGCCTCTGGAATTCCTCCCAATGAGTGGCCGAGGTGCCCAAGCTGTAGAGATTAGT	1379
QY	1347	caagaacacacacccaaagatttcaaccagcagctccaatatctgcgaagcaagtgta	1406
Db	1380	CAAGGACATCACACCAAAAGTTTTACCAAGGGAGGTCCGAATCTGCGMAAGCCACGGTA	1439
QY	1407	actgcacatgaaccgagagttgtgtcagggggacactgcgcccacagccgagacatt	1466
Db	1440	ACTGCCATGATAGCCCGAAGAGTTGTTGTAATGGGGCACCCTGCCCCACAGCCGAGACATT	1499
QY	1467	ltaaagatgatactcttccagggccacgtaccccaatggaacctgcacatagaccctctgag	1526
Db	1500	TTAAGAATAAACACCTCTCAGGCCACAGTACCCTGATGACCTCAGAGAACCTCTGAG	1559
QY	1527	caactgtactactcttccagagcccaaggatctccaggttgaaatacaagaatttcccaag	1586
Db	1560	CAACTGGACATATCTTTCCAAAGTCCAGGGATTTCCAGGTTGAATACAAAGACCTCCCAAA	1619
QY	1587	aacaaagaagaagagtgatattctcattcaactcigtctctcaacagcggcctctlogta	1646
Db	1620	AACACACAGAACGAATTTGATATCTTTATCAATGTGCTCTCCACACCACCTGTATGATACG	1679
QY	1647	catgacatcggaagaagtgtgagatcctgtcatalatagatgctgcacgtacaatttlaag	1706
Db	1680	CATGATTCGGCAAGAGATGTGAGTCTCTGCATGATATGCTGCGCCTGACATCTTAAAG	1739
QY	1707	ctgctgtcgtagttggaaccaacagacagatgccaagaatgacaagaagaatggaaccagt	1766
Db	1740	TTGCTGCTCATGATGTGACCAACAAATTAAGAGATGCCAAGAACAGGAAACGGACAATG	1799
QY	1767	tcaagctgaggagggtgtgtgaaccttcttgccacaacaacatttaaaa-cccaacata	1825
Db	1800	TCTGTGTGTGGAGGTCTGTAACCTTTTCTGGCAGTAAGCCCTTAAATATCCAAACTA	1859
QY	1826	tatactaaaaatactggaactgcttgaataattggaatactatataactcagtgagg	1885
Db	1860	TATACTGAATAATACTGA-AACTGCTTTGGAATAATTGGAATTTCTCATACCTCCAGTGGCG	1918
QY	1886	caagaca 1892	

Db	1919	CGAGAGA	1925	
RESULT	9			
LOCUS	HSAL32258	3066 bp	mRNA	PRI 14-MAY-1999
DEFINITION	Homo sapiens mRNA for stauflen protein, partial.			
ACCESSION	AJ132258			
VERSION	AJ132258.1	GI:4572587		
KEYWORDS	stauflen gene; stauflen protein.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 3066) Falcon,A.M., Fortes,P., Marion,R.M., Beloso,A. and Ortin,J.			
TITLE	Intersection of influenza virus NS1 protein and the human homologue of Stauflen in vivo and in vitro			
JOURNAL	Nucleic Acids Res. 27 (11), 2241-2247 (1999)			
MEDLINE	99263025			
REFERENCE	2 (bases 1 to 3066)			
AUTHORS	Ortin,J.			
TITLE	Direct Submission			
JOURNAL	Submitted (25-JUN-1999) Ortin J., Molecular and Cellular Biology, Centro Nacional de Biotecnologia, Cantoblanco, 28049 Madrid, SPAIN			
FEATURES	Location/Qualifiers			
source	1..3066 /organism="Homo sapiens" /db_xref="taxon:9606"			
gene	36..1526 /gene="stauflen"			
CDS	36..1526 /gene="stauflen" /function="dsRNA binding protein" /codon_start=1 /product="stauflen protein" /protein_id="CAB40082.1" /db_xref="GI:4572588"			
	translation="MKLGKPKMYKRPDVPYSKMSQTYNMMRGAVPRRYEYPFPVPL SEISQVEIHLKRLPVNFEPARFSGPKPMFPTKVSVEFGEGSKSKKKNNA LYQELASVGGQFNGKGTROAKDAAKRLRIQNEPILPERLEVNGRESEENLNK AIANVEELKRLPLPAYERVRYKPKRKTKTVKQTSPEYQGINPISRLAQIOAKK EKKEEVLTLERGLPRRREYPMOVKVNHNHAEQTYNKKAKRNAENMELTQKVP QRSHQTRQVAGEDTHKETRGWKSNNLMLALGEMGTSKDEFPMPYLSHQDLP GILDEPEVAQAVSQGHKTDFTRAAPNPAKATVAMLARLDELVGSTPTAETIK NNISGVHPGRLPRSEOLDLYSRVQGFQVEYEDFPKNNKNEFVSLINCSQPLIS HGICKDVESCHDMAALNLIKLSRLDOOSTFMPRTGNGPKMSVCGRC"			
BASE COUNT	891 a 693 c 699 g 783 t			
ORIGIN				
Query Match	42.3%; Score 1209.2; DB 68; Length 3066;			
Best Local Similarity	86.2%; Pred. No. 0; Mismatches 213; Indels 12; Gaps 6;			
Matches 1409; Conservative	0;			
268	gtagactctacccttaagctcaatgacgtctgtgtgaacgtggaagaacccaatgt	327		
1	GCATNACCCCTNACTAGACTAATGACACTGTGCATGAATCTTGCAAAAAACCAATGT	60		
328	ataagcccgtagaccctcaactctcgatgacgtacacctaacagctatgacgtgag	387		
61	ATAACCCGTGAGCCCTACTCTCGAGTGCAGTCCACCATTAACAAACATGAGAGGAG	120		
388	gtgactatccccccagataactcttaccatctccagctccagctccacttactcacaagt	447		
121	GTGCTTATCCCGGAGGACTTTTACCACTTCCAGTTCACCTTTACTTTATCAAGTGG	180		
448	agctctcgatgagcgagcaagcagcttaatlggaaagagaagatgagaccccgtagaac	507		
181	AACTTTCTGTGGAGACAGACGATTTAATGCGAAGAAAGCAAGACAGCGCTGGGAAC	240		
508	acgatgccctctgcccgtgcgtgaggaactctgcagagtgaaccctctgccagaaggttgg	567		

```

Db      241  ACGATGCTGCTGCAAGCGTTGAGGATCTGACAAATGAGCCCTGCGAGAGAGCTGG 300
Oy      568  aggtatattggaagagaagcagagagaagaacccaataatcgaataaacaagaagt 627
Db      301  AGTAAATGGAAGAAATTCGAAGAAATAATCTCAATAAATCTCAATAAATCTCAAT 360
Oy      628  ttgaattgagcttgagagcgaatttgcccttgaaatttgaggtggccggagagtg 687
Db      361  TTGACATTGCACTTAAAGGAACTTGCTGTGAATTTTGAGATTGCGCGGAGAGTGGC 420
Oy      668  caacacacatgaagaacttgtagcagggttcagttggggaatttgtagggagaag 747
Db      421  CACCCACATGAAGAACTTTGTGACCAAGTTTCGGTTGGGAGTTTGGGGGAAAGTG 480
Oy      748  aaggaaaaagcagaagatltccaagaagaatgcggccaggctgttcggagccgtta 807
Db      461  AAGGGAAGAAAGCAAGAAATTTCAAGAAAAATGCGCATACCTGTTCTTGAAGAGCTGA 540
Oy      808  ggaagctgcaccctccctcctgtgtgagcgagtgagcccgaaatcaagaagaagtc 867
Db      541  AGAATTACCGCCCTGCTGCTGAGTGAACGATTAAGCTTGAATCAAAAAAACAACA 600
Oy      868  agccacacatga-----agacagcccgagatltgccaagaggaatgaatcattagta 921
Db      601  AACCATATGTCAGGCAACAGACAGCAAGCAGAAATATGCGCAGGAGTCAATCCGATTAGCC 660
Oy      922  gacttgacacagatccagcagcgaagagaagagcagagatcagatgctcctcaag 981
Db      661  GACTGGCCCAAGATCCAGCAGCGCAAGAAAGAGAGAGAGCAAGATCAAGCTCCACAG 720
Oy      982  aacgagcttccacgctgcagcgagaggttgtgtagcagtgaaagttggagctacactg 1041
Db      721  ACGGAGGCTCTCCCGGCGCCGACAGGAGTTTGATGTCAGGTGAAGTTGAAACCACTG 780
Oy      1042  cagaagagtggtgtaaccaataaagaatggcgaagcgtatgctgtgagaaactgtg 1101
Db      781  CAGAAGAAAGGGGACCAACAAGAAAGGTGGCCAGCGCAATGACCGAGACATGTGG 840
Oy      1102  agatccctgggttcaaaagttccccaagcgagcagctgcgaagccagatcaataagaag 1161
Db      841  AGATCTTGGTTTCAAGTCCCGAGCGCAAGCA--CCAAACCCCGCACTCAAGTCAAGAG 898
Oy      1162  agaagaccctcagtaagaagaacagagagcagagaagaagtaa-cgttttttaacctcag 1220
Db      899  AGAAGACACCCCTTAAGAAACACAGGGGATGGAAGAAAGTAACTTTTGTGAACTGGC 958
Oy      1221  cctgggagat-gaaaaatggaatagtaacaagaagcagaggttcagagatgcttacttag 1279
Db      959  TCTGGGATGGAATAATGGGACTAGTAATAAGAGATGAGTTCAAGATGCCCTATCTAAG 1018
Oy      1280  ccatacgaagctgcagacttgaaatttcctcccaatggctccgaaggttgcccaagctgtcg 1339
Db      1019  TCATAGAGAGCTGCTGTGGAATTTCTCCATGTGTGCCCGAGGTGCCAGAGCTGTAG 1078
Oy      1340  ggttagtaagaagacacacacacacacacacacacacacacacacacacacacacac 1399
Db      1079  ACTTAGTCAAGGACATTCACACCAAAAGATTTTACCAAGGCGAGCTCCGAATCTGCAAGGC 1138
Oy      1400  aacgtaactgcatalagaccgagagatgtgtgtaaggggacactgcgccaacagcga 1459
Db      1139  CACGGTACTGCGCATGATAGGCCGAGAGTGTGTATGAGGGGACACTCGCCACAGCCGA 1198
Oy      1460  gacatctttaagaatgaacatctctcagcgacgtaccccaatgaaactgcagcagacc 1519
Db      1199  GACCATTTTAAAGATATCATCTTTCAGGGCCACGTACCCCATGACCTCTACACAGACC 1258
Oy      1520  cctctagaacactgactactccttcagagccagagattccaggttgaaatacaagatt 1579
Db      1239  CTCTAGACACATGACATCTTCTTCCAGAGTCCAGGAGATTCACAGTTGAATACAAAGACTT 1318
Oy      1580  tcccaagaacaacaagaacgagtgtagtctctcaatcaatgctcctcaacagcgctct 1639

```

```

Db      1319  CCCCCAAAACAAGAAAGAAATTTGTATCTCTATCAATGCTCTCTACACCACTCT 1378
Oy      1640  cgtcagcatgcatcggaagagatgtgagtcctgcatagatagatgctgacatgaat 1699
Db      1379  GATCAGCCATGATGATCGGCAAGATGTGAGTCTGCGCATGATATAGGCTGCCCTAAT 1438
Oy      1700  tttaagctgctgtctgagtttgagcaacaagagacagagatgccaagaagaagaatg 1759
Db      1439  CTTAAAGTTGCTGTCTGAGTTGACCAACAAAGTACAGAGATGCCAAGAACAGAAACGG 1498
Oy      1760  accagttcagcgtgcggagaggtgctgaaactcttctcgtgccaacaacatataaa-cc 1818
Db      1499  ACCAATGCTGTGTGGAGGTGCTGAACCTTTCTGCGCATGAACCATTAATAAATTC 1558
Oy      1819  caacatatactgaataatcagagactgtttgaaaattggataatcagaaactc 1878
Db      1559  CACATATATACTGAAATACTGA-AACTGCTTTGAAATTTGGAATTTGATATCTTC 1617
Oy      1879  agtgggccaagaca 1892
Db      1618  AGTGGCGGAGAGA 1631

```

```

RESULT 10
AC063967/c
LOCUS
DEFINITION Mus musculus clone RP23-41G23, WORKING DRAFT SEQUENCE, 13 unordered
pieces.
AC063967
VERSION AC063967.3 GI:9972323
HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE
AUTHORS McCombie, W.R., Baker, J.P., Bahret, A., Bai, H., Dedhia, N.N., de la
Bastide, M., Huang, E.N., King, L., Kirchhoff, K.A., Miller, B.,
Nascimento, L.U., O'Shaughnessy, A.L., Preston, R.R., Rodriguez, M.A.,
Shah, R.S., Shekher, M., Spiegel, L.A., Toth, K. and Vill, M.D.
TITLE Mouse Genomic Sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 225016)
AUTHORS McCombie, W.R.
TITLE Direct Submission
JOURNAL Submitted (22-APR-2000) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
On Sep 6, 2000 this sequence version replaced gi:8439872.
----- Genome Center
Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
Laboratory
Center code: CSHL
Web site: http://www.cshl.org/genseq
Contact: mcombie@cshl.org
----- Project Information
Center project name: RP23-41G23
Center clone name: RP23-41G23

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 121343: contig of 121343 bp in length
* 121344 121985: gap of unknown length
* 121986 139930: contig of 17945 bp in length
* 139931 140572: gap of unknown length
* 140573 155472: contig of 14900 bp in length
* 155473 156114: gap of unknown length

```

Query Match	Best Local Similarity	Matches 1083; Conservative	36.6%; Score 1045.2; DB 52; Length 225016; 99.0%; Pred. No. 3.8e-272;	0; Mismatches 8; Indels 3; Gaps 3;
OY 1766	ttcagcgtcgaggaagtgcgtgaacctttcttgcacaaacccatataaaccacaata	156115	167728; contig of 11614 bp in length	
Db 83313	TTTAAAGGGGCGGAGAGTGTGTGAACCTTTCTGCGCAAAACCATTAATAAAACCAACATA	157729	168370; gap of unknown length	
OY 1826	tatacttgaaataatcagagaaactgcctttgaaaaatttggaatactatcattactcagttggc	158371	177185; contig of 8815 bp in length	
Db 83253	TATACTGAAATAATCTGAGAGACTGCTTTGAAAAATTGGAAATATCTGATAATCTCCAGTGGGC	177186	177827; gap of unknown length	
OY 1886	caagacatcgtctgataataaatcttgcacaaagacacaaagaanaacttcagttgtagccctcg	177828	185945; contig of 8118 bp in length	
Db 83193	CAAGACATCTGTGTGATATAAATGTGGCAAAAGACAGACAAACTTCAGTGGTGAAGCCCTTG	185946	186587; gap of unknown length	
OY 1946	ttgtgcct-ggcgcgtagtgatgatactgctgtcgtcgtccatccatccagacagaacaccagcc	186588	194327; contig of 7740 bp in length	
Db 83133	TTGTGCTGGGCGCGGTAGTGATGATGCTGTGCTGTGCATCCATCCAGACAGAAACACAGCC	194328	194969; gap of unknown length	
OY 2005	ccaacgctccagctctgttttcttgatcgttgacaaagagagacagacagccaattctcatgtc	194970	202341; contig of 7372 bp in length	
Db 83073	CCAAAGCCCTCCGTTCTGTTTTGGATCGTGCACAAAGAGAGACAGACCAATTCCTATGCT	202342	202983; gap of unknown length	
OY 2065	gacctcttcagaacttctgaaaaaccgcgacacgcacacacagagagaccttaatacagcgc	202984	210036; contig of 7053 bp in length	
Db 83013	GGCTTCTTCAGTACTTTTGAAAAACCCCGACAGCCACACACAGAGAGCCCTTATAGGGGC	210037	210677; gap of unknown length	
OY 2125	ccgagacctaaccgacagagaaaagccgcagttgcttccactatgcatatcacttgcagc	210678	216586; contig of 5909 bp in length	
Db 82953	CCGGAGCTTAAACGACACAGAGAAAA-ACCAGTGGCTTCCACTACGACATGACGACTCAGC	216587	217227; gap of unknown length	
OY 2185	tcgcgcacacagtagaacaccactgtlaaacactgtttctcttc-agtttaatttttttccct	217228	220778; contig of 3551 bp in length	
Db 82894	TCGGCCACACCGTAGACACACTCTTAACCACTGCTTCTCTTCAAGTTTCATTTTTTTCCTT	220779	221419; gap of unknown length	
OY 2244	tgatgatatacaactataatttcattcagttcccttgatcgtgtcttactactaccataga	221420	223928; contig of 2509 bp in length	
Db 82834	TGATTGATACCACTATATATTTTTCATTTCATTTCCCTTAGTGTGCTTACTTAACCTAGGA	223929	224569; gap of unknown length	
OY 2304	gtttgaaactgtcagtcatactgtaacttgcgcaagatcatcacagcccggtttgggtgcatctg	224570	225016; contig of 447 bp in length.	
Db 82774	GTTTGAANAACGTCTCATGTATGTAACTGGCAAGATACACAGCCCGGTTGGGTGGCAATTG			
OY 2364	tgccctcgtgcttgctgtaaacagttcttgaaattacaaccaagaatccttgatccctgcgcc			

Db	82714	TCGCTCTGGCTTGGCTGAACAGTTCTTGGATTTACACAGAACTCTTGACTGCCGCC	82655
Qy	2424	ttgtataaatlygacagcttaggacctlttaaacctltgaatcaaaagatatgctcttt	2483
Db	82654	TTGTATAATTCAGACAGCTTAGACTTTTAAACTTAGATCAAAAGATATGCTCTTTT	82595
Qy	2484	aacctatltttaagagacagactttaaahagagccctgacctttaccatataacaga	2543
Db	82594	AACCTTATTTTTTAAGGACAGACTTTTAAATAAGAGCCCTGACCTTTTACCATTAATACAGA	82535
Qy	2544	atttgcaaaagagtglttctcttgagagagtgaccttltttaccacaactacagacat	2603
Db	82534	ATTTGTCAAAAGAGATGTTTCTTGAGGAGGTAGCTTTTTTTACACACACTACAGACAT	82475
Qy	2604	accttgaagcccaagaagactacagagcttgtygtccctlaagagggcccaataacgtcaatcc	2663
Db	82474	ACCTGTAGGCCCCAGAAAGACTACAGGCTGTGTGCCCTAGAGGGCCCAATATACATCAATTC	82415
Qy	2664	aacctctaagtcggggaaaagtgtaeaagtttcctgtygtgcacaggggagagca	2723
Db	82414	AACCTCTAAGTCGGGGAAAGGACAGGTTTCTGTGCTGTGTGCACAGGGAGGCA	82355
Qy	2724	gtcaagcggagccctggggagaagagcatctgtggtccctcagagagccctgtctcaacctgg	2783
Db	82354	GGTCAGCTGGCTGGGGGAAAGACATTTGGCTCTCAGTGCAGCCCTGCTTCACACTTGG	82295
Qy	2784	tttagctggaacctccacactcaatgaaatataagtaaacactcacttcttgtcacaaata	2843
Db	82294	TTTACCTGGAACCTTCCACATCATGGAATATATGAATAACTCACTTCTTTGCACCAATA	82235
Qy	2844	aatgtaataactaa 2857	
Db	82234	AATGGTAATACTAA 82221	
RESULT 11			
LOCUS	RNO010200		
DEFINITION	Rattus norvegicus mRNA for Staufen protein, partial.	22-JAN-1999	
ACCESSION	AJ010200		
VERSION	AJ010200.1 GI:4138427		
KEYWORDS	staufen gene.		
SOURCE	Norway rat.		
ORGANISM	Rattus norvegicus		
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;		
AUTHORS	Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
TITLE	1 (bases 1 to 360)		
JOURNAL	Kiebler, M.A.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (16-AUG-1998) Kiebler M.A., Cell Biology, EMBL,		
TITLE	Meyerhofstr. 1, D-69117 Heidelberg, GERMANY		
REFERENCE	2 (bases 1 to 360)		
AUTHORS	Kiebler, M.A., Hemraj, I., Verkade, P., Kohrmann, M., Fortes, P.,		
TITLE	Marton, R.M., Orlin, J. and Dotti, C.G.		
JOURNAL	The mammalian staufen protein localizes to the somatodendritic		
REFERENCE	domain of cultured hippocampal neurons: implications for its		
AUTHORS	involvement in mRNA transport		
TITLE	J. Neurosci. 19 (1), 288-297 (1999)		
JOURNAL	99080098		
MEDLINE			
FEATURES			
Source	Location/Qualifiers		
	1..360		
	/organism="Rattus norvegicus"		
	/db_xref="taxon:10116"		
	/clone_lib="Rat brain"		
	/dev_stage="embryonic"		
gene	1..119		
	/gene="staufen"		
	<1..>119		
CDs	/gene="staufen"		
	/codon_start=2		
	/evidence=experimental		
	/product="Staufen protein"		



```

/protein_id="CAA09037.1"
/db_xref="GI:4138428"
/translation="YFYPPVPPLLYQVELSVGGQDFNGKMKRPPVYKHDATA
BASE COUNT      109 a      70 c      105 g      76 t
ORIGIN

```

/db\_xref="GI:4138428"

```
/translation="YFYFPVPPLLYQVELSVGGQQENGKGMRRPVKHDATA"
```

BASE COUNT	109 a	70 c	105 g	76 t
ORIGIN				

Query Match	11.4%;	Score:325.4;	DB 74;	Length 360;
Best Local Similarity	94.2%;	Pred. NO. 4.5e-77;		
Matches 338; Conservative	0;	Mismatches 21;	Indels 0;	Gaps 0;

Oy	405	tactttaccatctccagctccacccttaactctaccacaagtctggcctccgtgggagaa	46
Db	2	TACTTTTACCATTTCCAGTTCCACCGTTACTCTACCAAGTTGAGCTCTCGTGGGAGGA	61
Oy	465	cagcagtttaatbvggnaagaaagatbgaaccacccgtbgaacaagatbccccctgc	52
Db	62	CAGCAGTTTAAATGGGAGAGGAAAGATGAGACCACTCTGAAACATGATGCCACTGCGCGA	12
Oy	525	ggcgctgaagactctgcagaagtgaacccctctgcagaagaagtgtgagatbtaatbgaagaa	58
Db	122	GCCCTGAGAGACTCTGCAGAGTGAAGCCCTTACCAGAGAGGCTTAGAGTGTAAATGGAAAGAA	18
Oy	585	gcagaagagaagaacccataaatcggnaataagaaacaaagtgttgtaaattgogctgaag	64
Db	182	TCAGAGAGAGAAACCTTAAATTAATTCAGAAATAAGCCAGAGTGTGATGGCCCTCAAG	24
Oy	645	cggaaattgacctbgaattcttgaagctggccgggagaaatgccccaccacataagaagac	70
Db	242	CGGAATTGGCTGTGAAATTTTGAGGTGGCCGGGAGAGATGGCCGCCACATATGAAGAC	30
Oy	705	tttgtgaccagtggttccagttgvggaaattctgaagggagaagagaagggaaaaagcaaga	76
Db	302	TTTGTGACCAAGGTTTCACTGTGGGAGAGTTTGTAGCGGAGAGGAAGGAAAAAGCAAGA	36

RESULT	12
AL133174/c	
LOCUS	113836 bp
DEFINITION	DNA
	PRI
AL133174	14-JUN-2000
Hunan DNA sequence from clone RP3-470L14 on chromosome 20, complete sequence.	

chromosome 0, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>

RP3-470L14 is from the library RPc1-3 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>

VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone RP3-470L14. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true right end of clone RP3-470L14 is at 11936 in this sequence. The true right end of clone RP1-1556 is at 100 in this sequence.

FEATURES	Location/Qualifiers
source	1. .113836

coverage. Assembly confirmed by restriction digest data.				
BASE COUNT	33974	a	25115	c
ORIGIN	24233	g	30514	t

Query Match	9.98;	Score 283.6;	DB 34;	Length 113836;
Best Local Similarity	92.58;	Pred. No. 1.7e-65;		
Matches 298;	Conservative	0;	Mismatches 24;	Indels 0;
			Gaps	0;

OY	1243	gtcaaaagacagagagatctcagatcagcttaacttagcacaacagagctgcagagttgaa	1302
Db	38168	GTAAATAAGAGAGATCAGTTCACAGATGCTTAACTAAGTCATCAGCAGCTCCTGCTGGAA	38109
OY	1303	ttctccccaatgctgcaggaagltgcacagctgtctggggtgaagttcaagaacacaccca	1362
Db	38108	TTCTTCCCATGATGGTGCCCGAGGTCGCCCAAGGCTGTAGGAGTTAGTAAAGACATTCACACA	38049
OY	1363	aagattccacagagcagctccaaatctctgcacaagcaacggttaactgcataatagccc	1422
Db	38048	AAGATTTCACAGGACACTCTCGAATTCCTGGCAAGGCCACGGTAACTGCAATGATAGCCC	37989
OY	1423	gaagatgtttgtacgggggacactctgcacacagccggaataccattttaaggtatacatct	1482
Db	37988	GAGATTTTGTGTATGGGGGCACTCTGCCACACGCCGAACCATTTTAAAGATTAACATCT	37929
OY	1483	cttcaagcgcacgtaccccatggaactctgcactatagaacctctgagcaactgtatacctt	1542
Db	37928	CTTCAGGCGCAGTACCCCATGGAACCTCTCAGAGAACCCCTGTAGCACTGGACATATCTTT	37869
OY	1543	ccagaagcccaaggtatccaggt	1564
Db	37868	CCAGAGTCCAGAGGATTCACAGT	37847

RESULT	13			
AK002152				
LOCUS	AK002152	2968 bp	mRNA	22-FEB-2000
DEFINITION	Homo sapiens CDNA FL11290 f15, clone pLACE1009622, weakly similar to MATERNAL EFFECT PROTEIN STAU6EN.		PRI	

ORGANISM	REFERENCE
Homo sapiens	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
Eutheria; Primates; Catarrhini; Hominoidea; Homo.	
1 (sites)	
Isoaga, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,	







repeat\_region 695. .3450  
/note="LIP3 repeat: matches 232. .2841 of consensus"  
repeat\_region 3451. .6427  
/note="LIP42 repeat: matches 3148. .6144 of consensus"  
repeat\_region 6432. .7143  
/note="LIM01 repeat: matches 5283. .6016 of consensus"  
repeat\_region 7164. .7316  
/note="Alub repeat: matches 137. .290 of consensus"  
repeat\_region 7339. .7646  
/note="LIMD2 repeat: matches 6016. .6330 of consensus"  
repeat\_region 7650. .7754  
/note="LIMC5 repeat: matches 7608. .7714 of consensus"  
repeat\_region 7767. .8072  
/note="Aluy repeat: matches 1. .305 of consensus"  
repeat\_region 8095. .8136  
/note="21 copies 2 mer ca 90% conserved"  
repeat\_region 8152. .8245  
/note="LIMC5 repeat: matches 7478. .7572 of consensus"  
repeat\_region 8273. .8494  
/note="Aluo repeat: matches 84. .305 of consensus"  
repeat\_region 8503. .8633  
/note="LIMC5 repeat: matches 7305. .7446 of consensus"  
repeat\_region 8712. .8879  
/note="84 copies 2 mer tt 72% conserved"  
repeat\_region 8899. .9190  
/note="Aluo repeat: matches 8. .300 of consensus"  
misc\_feature complement(9869. .10121)  
/note="match: STS: Em:LA1960. match: SNS: Em:LA1960"  
repeat\_region 12531. .13017  
/note="L2 repeat: matches 2083. .2505 of consensus"  
repeat\_region 13018. .13320  
/note="LIM1 repeat: matches 6004. .6303 of consensus"  
repeat\_region 13321. .14004  
/note="L2 repeat: matches 1197. .2083 of consensus"  
misc\_feature complement(16833. .17221)  
/note="match: GSS: Em:AQ558019"  
misc\_feature 16840. .17277  
/note="match: GSS: Em:B65742"  
misc\_feature complement(16845. .17241)  
/note="match: GSS: Em:AQ149727"  
misc\_feature complement(16848. .17220)  
/note="match: GSS: Em:AQ240745 Em:AQ474471"  
repeat\_region 16886. .17629  
/note="MER4D repeat: matches 234. .973 of consensus"  
repeat\_region 18183. .18407  
/note="LIP3 repeat: matches 5928. .6148 of consensus"  
repeat\_region 18455. .19002  
/note="MER41A repeat: matches 1. .554 of consensus"  
repeat\_region 19289. .19465  
/note="MIR repeat: matches 15. .198 of consensus"  
misc\_feature 19583. .20000  
/note="match: GSS: Em:AQ017861"  
repeat\_region 19875. .20274  
/note="MER57B repeat: matches 1. .403 of consensus"  
repeat\_region 20275. .21084  
/note="MER57-internal repeat: matches 1. .806 of consensus"  
repeat\_region 21078. .22422  
/note="MER57-internal repeat: matches 440. .2799 of consensus"  
repeat\_region 22425. .22764  
/note="MER57-internal repeat: matches 3892. .4230 of consensus"  
repeat\_region 22765. .23041  
/note="Alusc repeat: matches 8. .286 of consensus"  
repeat\_region 23042. .24919  
/note="MER57-internal repeat: matches 4230. .6114 of consensus"  
repeat\_region 24939. .25118  
/note="LIPB1 repeat: matches 5980. .6153 of consensus"  
repeat\_region 25119. .25422  
/note="Aluy repeat: matches 1. .304 of consensus"  
repeat\_region 25423. .25607  
/note="LIPB1 repeat: matches 5777. .5980 of consensus"

repeat\_region 25635. .25672  
/note="19 copies 2 mer ta 78% conserved"  
repeat\_region 25681. .26070  
/note="LIPB1 repeat: matches 5397. .5790 of consensus"  
repeat\_region 26082. .27136  
/note="MER57-internal repeat: matches 6489. .7537 of consensus"  
repeat\_region 27137. .27249  
/note="MER57B repeat: matches 1. .114 of consensus"  
repeat\_region 27250. .28097  
/note="LIMC3 repeat: matches 6763. .7631 of consensus"  
repeat\_region 28123. .28390  
/note="MTRIC repeat: matches 28. .313 of consensus"  
repeat\_region 28404. .28952  
/note="LIMC3 repeat: matches 5376. .5993 of consensus"  
repeat\_region 28946. .29622  
/note="LIM4 repeat: matches 4880. .5566 of consensus"  
repeat\_region 30851. .31163  
/note="Alusc repeat: matches 1. .301 of consensus"  
repeat\_region 32518. .34668  
/note="LIPB1 repeat: matches 3983. .6149 of consensus"  
repeat\_region 34672. .35189  
/note="MTR28 repeat: matches 1. .440 of consensus"  
repeat\_region 35190. .36499  
/note="LIPB1 repeat: matches 2718. .3988 of consensus"  
repeat\_region 36500. .36876  
/note="MTRIC repeat: matches 1. .369 of consensus"  
repeat\_region 36877. .38213  
/note="LIPB1 repeat: matches 1264. .2718 of consensus"  
repeat\_region 38221. .39169  
/note="LIPB1 repeat: matches 288. .1240 of consensus"  
repeat\_region 39455. .39517  
/note="L2 repeat: matches 2349. .2404 of consensus"  
repeat\_region 40251. .40337  
/note="MIR repeat: matches 59. .146 of consensus"  
repeat\_region 41354. .41648  
/note="Alusc repeat: matches 1. .295 of consensus"  
repeat\_region 41650. .41743  
/note="47 copies 2 mer aa 62% conserved"  
repeat\_region 41744. .42318  
/note="L1 repeat: matches 4553. .5128 of consensus"  
repeat\_region 42337. .42391  
/note="Alu repeat: matches 251. .302 of consensus"  
repeat\_region 42360. .42395  
/note="Alu repeat: matches 79. .114 of consensus"  
repeat\_region 42421. .43734  
/note="L1 repeat: matches 4019. .5334 of consensus"  
repeat\_region 43735. .43834  
/note="FLAM\_C repeat: matches 32. .131 of consensus"  
repeat\_region 43835. .44621  
/note="L1 repeat: matches 3232. .4019 of consensus"  
repeat\_region 44622. .44802  
/note="LIM1 repeat: matches 6121. .6304 of consensus"  
repeat\_region 44803. .45815  
/note="L1 repeat: matches 2126. .3232 of consensus"  
repeat\_region 45514. .46762  
/note="LIM1 repeat: matches 1155. .2794 of consensus"  
repeat\_region 46760. .47327  
/note="LIM1 repeat: matches 59. .642 of consensus"  
repeat\_region 47325. .48451  
/note="LIM1 repeat: matches -1334. .-154 of consensus"  
repeat\_region 48533. .48727  
/note="L1 repeat: matches 4030. .4226 of consensus"  
repeat\_region 48717. .49410  
/note="LIMB9 repeat: matches 5092. .5805 of consensus"  
repeat\_region 49411. .49506  
/note="48 copies 2 mer at 69% conserved"  
repeat\_region 49528. .50023  
/note="LIMB9 repeat: matches 5787. .6308 of consensus"  
repeat\_region 50050. .50333  
/note="Aluy repeat: matches 1. .300 of consensus"

Query Match

7.4%; Score 210.2; DB 68; Length 163253;

Best Local Similarity 60.1%; Pred. No. 1.4e-45;  
Matches 385; Conservative 0; Mismatches 253; Indels 3; Gaps 2;

```
OY 428 acccttactctacccaagtgaagctctccgttggcgagacagcgcttaatggaaagaaa 487
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159212 ACCTAAGATCTTTTATGTTCACGCTCATGTAGGAAATATGAAATTTTGGGGAAGGAAAG 159271

OY 488 gatgagaccaccgctgaaacacgatacgctcccttgcctgagacactctgcagatga 547
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159272 CCAGACACATCTGTGCAATGAAAGACATCCACAGCACTGCAGATTGACCTATTCCAGAAA 159331

OY 548 acccctgcagaaggttggagtaaatlygaagagaagcagaggaagaaacccaataa 607
      - - - - - ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159332 GATCCCTTAGAATGTGTAATCAGGAAAGCA--TATAGATGATGACAAAGATGCAAAATTA 159389

OY 608 atcggaataagccaagtggttgaatctgcgtgaagcgaaattgcctgtgaatttga 667
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159390 GCTGAGATCAGCTTAGTGTGTAATGCTGTGAATCTGAACTGAAATATACCTGTAGTTCTGA 159449

OY 668 ggtggcccgaggagtggtgcccacacacatagaagaacttltgaccaaggttcagttg 727
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159450 GGTATTAAAGAAAGTGGACCCACATATGAAAGCTTTGTACTCGAGTGTCAGTAGG 159509

OY 728 ggaatttgaaggaggaaggaaggaagcaagaagatctccaagaagaatgcggccag 787
      - - - - - ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159510 AAAGTTCTCTACAGAGAGAGGAAGAAATAGCAAAAAACTCTCCAAGAAAGCGTGTGCAAC 159569

OY 788 ggcgttctggagcagcttaagagagctccaccctcctgcgtgtgagcgagtgaagcc 847
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159570 CACCGTCTTACAGAGCTTAAAAAACTTCCACCTCTCTCTGTGTGTGAAAGGCAAAATTT 159629

OY 848 cagaaacaagaagaaagtcagcccaactgcagaacagcccccgaatlaltgccaaggaat 907
      - - - - - ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159630 ATTTTAAAAAAAGCCCTTAAACAAATAGTCAAGGCCGAGCCAGAAATA-GGCCAAGGGAT 159688

OY 908 gaatccatattagtagaactgcagacatccagcaggaacaaaggaagagagcagaagta 967
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159689 GAACCTTATTACCTGCTGCGCCCAATTCAACAGGCGCAAAAAGGAAAGGACGACGATTA 159748

OY 968 catgctccttaagaagagagctctccacgctgcagggaggttltgatacagtaagta 1027
      - - - - - ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159749 TGTTTTGTCTTCAGAAAGAGAAATGCTGATGTGAGAAATTTGTGATGTATATGAAAGGT 159808

OY 1028 tgggcatacaactgcagaagaggtgtgtacataaagaag 1068
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159809 AGGCAATGAGCTGCTACAGAGAGGAGACTAATAAAAAAG 159849
```

Search completed: April 5, 2001, 09:36:07  
Job time: 81412 sec



alignment block:

US-09-316-048-6 x 293322

Align seg 1/1 to: 293322 from: 1 to: 3506

1 MetSerGlnValGlnValGlnValGlnAsnProSerAlaAlaLeuSerG 17  
409 ATGTCGAAGTTCACAGTTCAGAACCCATCTGCTCTCTCAGG 458  
17 ySerGlnIleLeuAsnLysAsnGlnSerLeuLeuSerGlnProLeuMetS 34  
459 GAGCCAAATACGTGACAGAACACAGCTCTCTCTCAGACCTTGATGA 508  
34 erLleProSerThrThrSerSerLeuProSerGlnAsnIleGlyArgPro 50  
509 GTATTCCTTCTACTACTAGCTCTGCTGCTGAAATGAGAGGTACACC 558  
51 IleGlnAsnSerAlaLeuProSerAlaSerIleThrSerThrSerAlaI 67  
559 ATTCAAACTCTGCTTACCCTCTGCATCTATTACATCCACCGTGCAGC 608  
67 alaGluSerIleThrProThrValGlnLeuAsnAlaLeuCysMetLysL 84  
609 TGCAGAAACCATTAACCCCTACTAGAACTAAATGCACGTGCATGAAC 658  
84 euGlyLysLysProMetLysProLysProValAspProThrSerArgMetGln 100  
659 TTGGAAAAAACCAATGTAATAGCCGTGGACCTTACTCTCGATGACAG 708  
101 SerThrTyraAsnTyraAsnMetArgGlyGlyAlaTyraProProArgTyraPh 117  
709 TCCACCTATAACTACACACATGAGAGAGAGGCTTATCCCCGGAGTACTT 758  
117 eTyraProPheProValProProLeuLeuTyraGlnValGlnLeuSerValG 134  
759 TTACCCATTTCCAGTTCACCTTACTTATCAAGTGAACTTTCTGTGG 808  
134 yGlyGlnGlnPheAsnGlyLysGlyLysThrArgGlnIleAlaLysHis 150  
809 GAGGACAGCAATTTAATGCAAGAAAGCAAGACAGCGCTGCGAAACAC 858  
151 AspAlaAlaAlaLysAlaLeuArgIleLeuGlnAsnGlnProLeuProG 167  
859 GATGCTGCTGCCAAACGCTTGAAGATCTGCAGAAATGAGCCCTGCCACA 908  
167 uArgLeuGlnValAsnGlyArgGlnSerGlnGlnIleAsnLeuAsnLys 184  
909 GAGGCTGAGGTGAATGAAAGAAATCCGAAAGAAATCTCAATAAAT 958  
184 erGlnIleSerGlnValPheGlnIleAlaLeuLysArgAsnLeuProVal 200  
959 CTGAAATTAAGTCAAGTGTGAGATTGCACTTAACGGAACTTGCCGTGG 1008  
201 AsnPheGlnValAlaIleArgGlnSerGlyProProHisMetLysAsnPhe 217  
1009 AATTTGAGAGTGGCCCGGAGAGTGGCCCAACCCACATGAGAGACTTTGT 1058  
217 lThrLysValSerValGlyLupPheValGlyLugLugLugLysSerL 234  
1059 GACCAAGGTTTGGTGGGAGTTTGTGGGGGAAAGTGAAGGGAAGCA 1108  
234 yLysIleSerLysLysAsnAlaIleAlaValLeuGlnGlnLeuLys 250  
1109 AGAAGATTTCAAGAAAGAAATGCCGCATAGCTGCTTGAGGAGCTGAAG 1158  
251 LysLeuProProLeuProAlaValGlnArgValLysProArgIleLysL 267  
1159 AAGTTACCGCCCTGCTGAGTTGAACGATGAAGCTTAGAATTCAAAAA 1208  
267 sLysThrLysProIleValLysProGlnThrSerProGluTyraGlyGln 284  
1209 GAAAGCAAAACCATGATGTCAGGCACAGCAAGCCCAAGATATATGGCAGG 1258

284 yLLeAsnProIleSerArgLeuAlaGlnIleGlnGlnAlaLysGln 300  
1259 GATCAATCCGATTACCGACAGTGGCCAGATCCAGCAAGCAAAAAAGGAG 1308  
301 LysGluProGluTyraThrLeuLeuThrGluArgGlyLeuProArgArgAr 317  
1309 AAGGAGCCAGAGTACAGCGCTCTCACAGAGCGAGGCTCCCGCGCCGAG 1358  
317 gGluPheValMetGlnValLysValGlyAsnHisThrAlaGluGlyThrG 334  
1359 GGAGTTGTGATGACAGTGAAGTTGAAACCACTGCAAGAGGAACGG 1408  
334 yThrAsnLysLysValAlaLysArgAsnAlaIleGlnAsnMetLeuGln 350  
1409 GCACCAACAAAGAGGTGGCCAGACGCATGACCCAGAAACATGCTGGAG 1458  
351 IleLeuGlyPheLysValProGlnArgGlnProThrLysProAlaLeuL 367  
1459 ATCCTTGTTCAAGTCCCGCAGCGGAGCCACCAACCCCGCAGTCAA 1508  
367 sSerGluGluLysThrProIleLysLysProGlyAspGlyArgLysValT 384  
1509 GTCAGAGAGAAAGACACCCATAAAGAACCGAGGATGGAAGAAAGTAA 1558  
384 hrPhePheGluProGlySerGlyAspGlnAsnGlyThrSerAsnLysGlu 400  
1559 CCTTTTGAACCTGGCTCTGGGATGMAAATGGAGTAAATTAAGAG 1608  
401 AspGluPheArgMetProTyraLeuSerHisGlnGlnLeuProAlaGlyI 417  
1609 GATGAGTTCAGATGCTTATTAAGTACATCAGCAGCTGCTCTGGAAT 1658  
417 eLeuProMetValProGluValAlaGlnAlaValGlyValSerGlnGlyH 434  
1659 TCTTCCATGTGTCGCCAGAGTCCGCCAGGCTGAGAGATTAGCAAGGAC 1708  
434 lHisThrLysAspPheThrArgAlaAlaProAsnProAlaLysAlaThr 450  
1709 ATCACACCAAGATTTTACCAGGGCAGCTCGAATCTCGCAAGGCCACG 1758  
451 ValThrAlaMetIleAlaIleArgGlnLeuLeuTyraGlyLysThrSerProTh 467  
1759 GTAACGTCCATGATAGCCCGAGAGTTGTATGGGGCCACCTCGCCAC 1808  
467 lAlaGluThrIleLeuLysAsnAsnIleSerSerGlyHisValProHisG 484  
1809 AGCCGAGACCATTTTAAGATTAACATCTTTCAGGCCACAGTACCCATG 1858  
484 yProLeuThrArgProSerGlnGlnLeuAspTyraLeuSerArgValGln 500  
1859 GACCTCTCAGAGACCCCTCTGAGCAACTGGACTATCTTTCCAGAGTCCAG 1908  
501 GlyPheGlnValGluTyraLysAspPheProLysAsnAsnLysAsnLuph 517  
1909 GATTCGAGTTGAATACAAAGACTTCCCAAAAACAAACAGAACGAATT 1958  
517 eValSerLeuIleAsnCysSerSerGlnProProLeuIleSerHisGlyI 534  
1959 TGTATCTTATCAATTTGCTCTCAGCCACCTCGATCAGCCAGTGA 2008  
534 lGlyLysAspValGlnSerCysHisAspMetAlaIleLeuAsnIleLeu 550  
2009 TCGGCAAGAGTGTGAGTCTTCGATGATATGGCTCGGTGAACATCTTA 2058  
551 LysLeuLeuSerGlnLeuAspGlnGlnSerThrGlnMetProArgThrG 567  
2059 AAGTTCTGTCTGATGTTGAGCAACAAAGTACAGAAATGCAAGAACAGG 2108  
567 yAsnGlyProMetSerValCysGlyArgCys 577  
2109 AAGCGACCAATGTGTGTGTGGAGGTGC 2139

seq\_name: /SID6/gcgdata/geneseq/geneseqn/NA2000.DAT.Z93324

seq\_documentation\_block:

ID Z93324 standard; cDNA: 3142 BP.

AC Z93324;

DT 04-JUL-2000 (first entry)

DE Human staufen cDNA.

XX Staufen: RNA binding protein; RNA/DNA hybrid; virus; retrovirus;  
 KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;  
 KW acquired immune deficiency syndrome; protease; human; ss.

XX Homo sapiens.

PH Key Location/Qualifiers

FT CDS 288..1193

FT /tag= a

XX /product= Staufen protein

PN CA2238656-A1.

PD 22-NOV-1999.

XX 22-MAY-1998; 98CA-2238656.

XX 22-MAY-1998; 98CA-2238656.

XX (UYMO-) UNIV MONTREAL.

XX Duchaine T, Wickham L, Cohen EA, Luo M, Desgroselliers L;

PI Moulard AJ;

XX WPI: 2000-246924/22.

DR P-PSDB: Y83108.

XX Isolated mammalian or Caenorhabditis elegans staufen proteins useful

PT for treating retroviral infections especially human immunodeficiency

PT virus infections (HIV)

XX Claim 4; Figure 1b; 96pp; English.

XX Staufen is a RNA binding protein which interacts with double stranded

CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA

CC transport and localization. Mammalian SP contain multiple ds

CC RNA-binding domains. Chimeric proteins comprising portions of

CC staufen with proteins with RNase or protease activity can be used to

CC treat virus infections. The RNase or protease activity of the fusion

CC protein prevents proper maturation of the virus. Mammalian staufen

CC proteins seem to recognise double stranded RNA structure rather than

CC any sequence specific position. 2 bacterially-expressed fusion

CC proteins used in an RNA-binding assay (his/hsp (human SP) and

CC MBP/MSP (murine SP)) both proteins strongly bound double stranded

CC RNA. Both fusion proteins also directly bound labeled ds RNAs and

CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the

CC treatment of RNA virus infections, especially those caused by

CC retroviruses, in particular human immunodeficiency virus. This

CC sequence is an alternatively spliced sequence to the one given in

XX Z93322.

XX Sequence 3142 BP; 862 A; 778 C; 733 G; 769 T; 0 other;

alignment\_scores:

Quality: 2664.50

Ratio: 4.990

Percent Similarity: 94.849

Percent Identity: 93.073

alignment\_block:

US-09-316-048-6 x Z93324

Align seq 1/1 to: Z93324 from: 1 to: 3142

```

15 LeuSerglySerglnIleLeuAsnLysaNGlnSerLeuSerglnPr 31
   ||| ||| :|: :|: ||| ||| |||
121 CTCACGCGCCACATCGGCTCTTCCCTCCCTTCGTCCTTCTTCTCC 170
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
31 OleuMetSerIleProSerThrThrSerSerLeuProSerglnaIag 48
   :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
171 TTTTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 219
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
48 lYargProIleGlnAsnSerAlaLeuProSerAlaSerIleThrSerThr 64
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
220 GCCCGGGG.....ACGAGTCGC 236
65 SerAlaAlaIagIleSerIleThrProThrValGlnLeuAsnAlaLeuCy 81
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
237 GAGCAGCAGCCAGAAAGCATACCCCTACTGTAGACTAAATGCACGTGTG 286
81 smetLysLeuGlyLysLysPrometYrLysProValAspProTyrSera 98
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
287 CATGAACCTTGGAAGAAACCAATGTATAGCCTGTGACCCCTTACTCTC 336
98 rGmetGlnSerThrTyrAsnTyrAsnMetLargGlyAlaIatYrProPro 114
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
337 GGATGCAGTCCACCTATTAACAACATGAGAGAGAGGTCTTATCCCGC 386
115 ArgTyrPheTyrProPheProValProProLeuLeuTyrGlnValGlnLe 131
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
387 AGGTACTTTACCCATTTCAGTTCACCTTACTTATCAAGTGGAACT 436
131 uSerValGlyGlnGlnPheAsnGlyLysGlyLysThrArgGlnAla 148
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
437 TTTCTGTGGAGAGACAGCAATTAATGGCAAGAAAGCAAGACAGCGTGT 486
148 lAlYshIsAspAlaAlaAlaLysAlaLeuArgIleLeuGlnAsnGlnPro 164
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
487 CGAAACACGATGCTGCTGCCAAGAGGTGAGAGATCTCCAGAAATGAGCC 536
165 LeuProGluArgLeuGluValAsnGlyArgGlnSerGlnGlnGlnLeu 181
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
537 CTGCCAGAGAGGCTGGAGGTGATGAGAGATGAGATGAGATGAGATGAG 586
181 uAsnLysSerglnIleSerglnValPheGlnIleAlaLeuLysArgAsnL 198
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
587 CAATTAATCTGAATTAAGTCAAGTGTGAGATGCACTTAACGAGAACT 636
198 euProValAsnPheGlnValAlaArgGlnSerGlyProProHisMetLys 214
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
637 TGCCCTGTGAATTTGAGAGTGGCCCGGAGAGTGGCCACCCACATGAA 686
215 AsnPheValThrLysValSerValGlyGlnPheValGlyGlnGlnGln 231
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
687 AACTTTGTGACCAAGGTTTCGTTGGGAGTTTGTGGGGAAGGTGAAGG 736
231 YLysSerLysLysLysLysLysLysLysLysLysLysLysLysLysLys 248
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
737 GAAAGACAGAAAGATTTCAAGAAAGAAAGTCCGCACTACTCTGTGAGG 786
248 luleuLysLysLeuProProLeuProAlaValGluArgValLysProArg 264
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
787 AGCTGAAGAAAGTTACCGCCCTGCTGCAGTTGAACGAGTAAAGCTAGA 836
265 lIleLysLysLysThrLysProIleValLysProGlnThrSerProGlnT 281
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
837 ATCAAAAGAAAGAAACCAATAGTCAAGCCACAGACAGCCAGAAATTA 886
281 rGlyGlnGlyIleAsnProIleSerArgLeuAlaGlnIleGlnAlaL 298
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
887 TGCCAGGAGATCAATCGATTAGCCGACTGCGCCAGATCCAGCAGGCAA 936
298 ySLysGluLysGluProGluTyrThrLeuLeuThrGluArgGlyLeuPro 314
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
937 AAAAGAGAGAGAGCCAGATACAGGCTCTCACAGAGCGAGGCTCCCG 986

```

```

315 ArgArgGluPheValMetGlnValIleGlyAsnHisThrAlaG 331
    |||||||
987 CGCGCGAGGAGGAGTTGTGATGACGAGTGAAGCTTGGAACACCACTGCAGA 1036
    |||||||
331 uGlyThrGlyThrAsnLysLysValAlaLysArgAsnAlaGluAsnM 348
    |||||||
1037 AGGAGCGGCGCACCAACAGAAAGTGGCCAGCCCAATGACGCCAGAACAA 1086
    |||||||
348 etLeuGluIleLeuGlyPheLysValProGlnArgGlnProThrLysPro 364
    |||||||
1087 TGTGTGGAGATCTTGTGTTCAAGTCCCGCAGCGGCGCCACCAAAACC 1136
    |||||||
365 AlaLeuLysSerGlnGluLysThrProIleLysLysProGlyAspGlyArg 381
    |||||||
1137 GCACCTACAGTCAGAGAGAGAAAGACACCATTAAGAAACAGGAGATGGAAG 1186
    |||||||
381 gLyValThrPhePheGluProGlySerGlyAspGluAsnGlyThrSerA 398
    |||||||
1187 AAAAGTAACCTTTTGTGAACCTGGCTCTGGGATGAATAATGGGACTAGTA 1236
    |||||||
398 snLysGluAspGluPheArgMetProTyrIleuSerHisGlnGlnLeuPro 414
    |||||||
1237 ATAAGAGGATGAGTTCAGATGCTTATCTAAGTCATCAGCAGCTGCTCT 1286
    |||||||
415 AlaGlyIleLeuProMetValProGluValAlaGlnAlaValGlyLys 431
    |||||||
1287 GCTGGATTTTCCCATGTCGTCGAGGTGCGCCAGGCTGAGAGAGTTAG 1336
    |||||||
431 rGlnGlyHisHisThrLysAspPheThrArgAlaAlaProAsnProAlaL 448
    |||||||
1337 TCAAGGACATCACACCAAGATTTTACACAGGCGACGCTCCAAATCTCTCCA 1386
    |||||||
448 ysaLathrValThrAlaMetIleAlaArgGluLeuLeuTyrGlyGlyThr 464
    |||||||
1387 AGGCCAGCGTAACCTGCATGATGACCCGAGAGTGTGTGAGGGGCGACCC 1436
    |||||||
465 SerProThrAlaGluThrIleLeuLysAsnAsnIleSerSerGlyHisVal 481
    |||||||
1437 TCCGCCACACCGCAGACCATTTTAAAGATTAACATCTCTTCCAGGCCAGT 1486
    |||||||
481 lProHisGlyProLeuThrArgProSerGlnGlnLeuAspTyrIleuSerA 498
    |||||||
1487 ACCCATGAGACTCTCAGAGACCCCTCTGAGCAACTGAGCATCTTCTTCCA 1536
    |||||||
498 rGlyAlaGlnGlyPheGlnValGluTyrLysAspPheProLysAsnAsnLys 514
    |||||||
1537 GAGTCCAGGAGATTCAGGTTGAATACAAAGACTTCCCAAAAACAACAG 1586
    |||||||
515 AsnGluPheValSerLeuIleAsnGlySerSerGlnProProLeuIleSe 531
    |||||||
1587 AACGAATTTTATCTCTTATCAATTTGCTCTCTCAGCCACCTCTGATCAG 1636
    |||||||
531 rHisGlyIleGlyLysAspValGluSerGlyHisAspMetAlaAlaLeuA 548
    |||||||
1637 CCATGGATATGCGCAGAGATGTGAGTCCATGATGATGCGCGCTGA 1686
    |||||||
548 snIleLeuLysLeuLeuSerGlnLeuAspGlnGlnIleSerThrGlnMetPro 564
    |||||||
1687 ACATCTTAAAGTGTGCTGTGAGTTCAGACCAACAAAGTACAGAGATGCCA 1736
    |||||||
565 ArgThrGlyAsnGlyProMetSerValGlyGlyArgGly 577
    |||||||
1737 AGAACAGGAAACGACCAATGTCTGTCTGTGGAGGTGC 1775
    |||||||

```

seq\_name: /SIDS6/gcgdata/geneseq/geneseqn/NA2000.DAT:293323

seq\_documentation\_block:

ID 293323 standard; cDNA; 3217 BP.

AC 293323;

DT 04-JUL-2000 (first entry)

```

XX XX Human staufen cDNA.
DE DE
XX XX
KW KW Staufen: RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW KW acquired immune deficiency syndrome; protease; human; ss.
OS OS Homo sapiens.
XX XX
FH FH Key Location/Qualifiers
FT FT CDS 363..1853
FT FT /*tag= a
FT FT /product= Staufen
XX XX
XX XX CA2238656-A1.
PD PD 22-NOV-1999.
XX XX
XX XX 22-MAY-1998; 98CA-2238656.
XX XX
XX XX 22-MAY-1998; 98CA-2238656.
XX XX
XX XX (UYMO-) UNIV MONTREAL.
XX XX
XX XX Duchaine T, Wickham L, Cohen EA, Luo M, Desrosiers L;
XX XX Moulard AJ;
XX XX
XX XX WPT: 2000-246924/22.
XX XX
XX XX P-PSDB: Y83108.
XX XX
XX XX Isolated mammalian or Caenorhabditis elegans staufen proteins useful
XX XX for treating retroviral infections especially human immunodeficiency
XX XX virus infections (HIV)
XX XX
XX XX Claim 4; Figure 1a; 96pp; English.
XX XX
XX XX Staufen is a RNA binding protein which interacts with double stranded
XX XX RNA and/or rough endoplasmic reticulum. It is involved in mRNA
XX XX transport and localization. Mammalian SP contain multiple ds
XX XX RNA-binding domains. Chimeric proteins comprising portions of
XX XX Staufen with proteins with RNase or protease activity can be used to
XX XX treat virus infections. The RNase or protease activity of the fusion
XX XX protein prevents proper maturation of the virus. Mammalian staufen
XX XX proteins seem to recognise double stranded RNA structure rather than
XX XX any sequence specific position. 2 bacterially-expressed fusion
XX XX proteins used in an RNA-binding assay (his/NSP (human SP) and
XX XX MBP/MSF (murine SP)) both proteins strongly bound double stranded
XX XX RNA. Both fusion proteins also directly bound labeled ds RNAs and
XX XX RNA/DNA hybrids. Staufen proteins may therefore be useful in the
XX XX treatment of RNA virus infections, especially those caused by
XX XX retroviruses, in particular human immunodeficiency virus. This
XX XX sequence is an alternatively spliced sequence to the one given in
XX XX 293322.
XX XX
XX XX Seq. Sequence 3217 BP; 889 A; 796 C; 742 G; 790 T; 0 other;
XX XX

```

alignment\_scores:  
 Quality: 2659.00 Length: 577  
 Ratio: 4.906 Gaps: 3  
 Percent Similarity: 93.934 Percent Identity: 91.508

alignment\_block:  
 US-09-316-048-6 x 293323 ..

Align seg 1/1 to: 293323 from: 1 to: 3217

```

15 LeuSerGlySerGlnIleLeuAsnLysAsnGlnSerLeuLeuSerGlnPr 31
    ||| |||
121 CTCACGCGCCACCTCCGCTCTTCCCTCCCTCTCTCTCTCTCTCTCC 170
    ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
31 oLeuMetSerIleProSerThrThrSerSerLeuProSerGluAsnAlaG 48
    ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```





KW ovary; brain; heart; telomerase RNA; complex; catalytic subunit;  
 KW telomerase assembly; maturation; transport; regulation; interaction;  
 KW inhibitory agent; antibody; immunoassay; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 345..1835  
 FT /tag= a  
 FT /product= "Human staufen protein"  
 FT /note= "Binds to telomerase RNA"  
 FT 3172..3177  
 FT /tag= b  
 FT polyA\_signal  
 XX  
 PN M09951255-A1.  
 XX  
 PD 14-OCT-1999.  
 XX  
 PF 06-APR-1999; 99WO-US07533.  
 XX  
 PR 06-APR-1998; 98US-0080783.  
 XX  
 PA (UUYO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.  
 PI Greider CW, Le S;  
 XX  
 PI MPI: 1999-620168/53.  
 DR P-PSDB; Y26335.  
 DR  
 XX  
 PT Human staufen polypeptide useful in methods for identifying telomerase  
 PT inhibitors -  
 XX  
 PS Claim 13; Page 41-42; 50pp; English.  
 CC The present sequence is an isolated DNA encoding human staufen (hstauf)  
 CC protein. The hstauf protein was detected in protein lysates from lung,  
 CC kidney, testis and ovary, but not in brain and heart. It binds  
 CC specifically to human telomerase RNA and can complex with the  
 CC telomerase catalytic subunit. The hstauf protein may play a role in  
 CC telomerase assembly, maturation, transport and regulation. Interaction  
 CC between hstauf and telomerase can be inhibited by administering inhibitory  
 CC agents to the cell. These telomerase inhibitors can be identified using  
 CC hstauf protein. Antibodies generated against hstauf can be used in various  
 CC immunoassays.  
 CC  
 SO Sequence 3190 BP; 880 A; 788 C; 736 G; 786 T; 0 other;

alignment\_scores:  
 Quality: 2650.00 Length: 577  
 Ratio: 4.907 Gaps: 2  
 Percent Similarity: 93.588 Percent Identity: 90.815

alignment\_block:  
 US-09-316-048-6 x X90786 ..

Align seg 1/1 to: X90786 from: 1 to: 3190

15 LeuSerGlySerGlnIleLeuAsnLysAsnGlnSerLeuSerGlnPr 31  
 ||| ||| ::::|  
 103 CTCACAGGCGCACATCCCTTCCTCCCTCCCTTCCTTCCTTCCTCC 152  
 31 oueMetSerIleProSerThrThrSerSerLeuProSerGluAsnAlaG 48  
 |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
 153 TTTTTCCTTCCTTCCTTCCTCCCTTCCTGCGCGCACCGCCAGGACGCG 201  
 48 |yArGProIleGlnAsnSerAlaLeuPro..... 57  
 ||||| ::::|:::|  
 202 GCGGGGAGAGACGTCGAGACACAGCAGAGATTATTAACCACTTAAC 251  
 58 SerAlaSerIleThrSerThrSerAla.....Al 67  
 ||| ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
 252 TCTCAGAACTGAACAAAGACACATTTGTTCTGGAACGCCCTCTTTTAA 301

67 aAlaGluSerIleThrProThrValGluLeuAsnAlaLeuCysMetLysL 84  
 |||||  
 302 AAAGAAGACATACCCCTTACTAGACTAAATGCACATGTGATTAAC 351  
 84 euGlyLysLysProMetLysLysProValAspProThrSerArgMetGln 100  
 |||||  
 352 TTGGAAAAAACCAATGTATTAACCTGTGTGACCTTACTCTCGATGCGG 401  
 101 SerThrTyrAsnTyrAsnMetLArgGlyAlaTyrProProArgLysPr 117  
 |||||  
 402 TCCACCTATAACTACAAACATGAGAGAGGTGCTTATCCCGGAGTACT 451  
 117 eYrProPheProValProProLeuLeuTyrGlnValGluLeuSerValG 134  
 |||||  
 452 TTACCCATTTCCAGTTCACACTTACTTTATCAAGTGGAACCTTCTGTG 501  
 134 LysGlyGlnGlnPheAsnGlyLysGlyLysThrArgGlnAlaAlaLys 150  
 |||||  
 502 GAGGACAGCAATTTAATGGCAAGCAAGACAGACGCTCGCAAAAC 551  
 151 AspAlaAlaAlaLysAlaLeuArgIleLeuGlnAsnGluProLeuProG 167  
 |||||  
 552 GATGCTGCTGCCAAAGCGTTGAGGATCTGCAGAAATGAGCCCTGCCAG 601  
 167 uArgLeuGluValAsnGlyArgGluSerGluGluAsnLeuAsnLys 184  
 |||||  
 602 GAGGCTGAGAGTAAATGGAAGAAATCCGAGAAAGAAATCTCAATTAAT 651  
 184 eArgLysSerGlnValPheGluIleAlaLeuLysArgAsnLeuProVal 200  
 |||||  
 652 CTGAATAAGTCAAGTGTGAGATTCACCTTAACGGAACCTTGCTGTG 701  
 201 AsnPheGluValAlaArgGluSerGlyProProIleMetLysAsnPhe 217  
 |||||  
 702 AATTTGAGGTGGCCGGAGAGTGGCCACCCACATGTAAGACTTTGT 751  
 217 IThrLysValSerValGlyGluPheValGlyGluGlyLysSerL 234  
 |||||  
 752 GACCAAGTTCGTTGGGAGTGTGGGGAGAGTGAAGGAAAGCA 801  
 234 LysLysIleSerLysLysAsnAlaAlaIleAlaValLeuGluGluLeu 250  
 |||||  
 802 AGAAGATTTCAAAGAAAAATGCCGCATAGCTGTGTTGAGGAGCTGAAG 851  
 251 LysLeuProProLeuProAlaValGluArgValLysProArgLysLys 267  
 |||||  
 852 AAGTTACCGCCCTGCTGCAGTTGAACGATGAACCTTAAGATCAAAA 901  
 267 sLysThrLysProIleValLysProGlnThrSerProGluTyrGlyGln 284  
 |||||  
 902 GAACAACAAACCATATGTCACAGCACAGCAACCCAGATATGCGCAG 951  
 284 LysLeuAsnProIleSerArgLeuAlaGlnIleGlnAlaLysLysGlu 300  
 |||||  
 952 GGATCAATCCGATTACCGCAGCTGCGCAGATCCAGCAGCGCAAAAAGGAG 1001  
 301 LysGluProGluTyrThrLeuLeuThrGluArgGlyLeuProArgLys 317  
 |||||  
 1002 AAGGAGCCAGAGTACAGGCTCTCACAGAGCGAGGCTCCGCGCGCAG 1051  
 317 gGluPheValMetGlnValLysValGlyAsnIleThrAlaGluLys 334  
 |||||  
 1052 GAGGTTTGATGATCAGTGAAGTTTGAAACCACTGCGAAGAGAACGG 1101  
 334 LysThrAsnLysLysValAlaLysArgAsnAlaIleGluAsnMetLeu 350  
 |||||  
 1102 GCACAAACAAAGAGTGGCCAAACCGCAATGCAAGCAATGCTGGAG 1151  
 351 IleLeuGlyPheLysValProGlnArgGlnProThrLysProAlaLeu 367  
 |||||  
 1152 ATCCTTGTTTCAAAGTCCCGCAGGCGCAGCCACCAACCCGCTACTCA 1201

```

367 sSerGIuLlLysThrProIleuLysLysProGIyAspGIyArGIyValT 384
|||||
1202 GTCAGAGAGAGACACCATTAAGAAACAGGAGATGGAAGAAAGTAA 1251
384 hrhepHeGluProGIySerGIyAspGIuAsnGIyThrSerAsnLysGIu 400
|||||
1252 CTTTTTTGAACCTGGCTGGGGATGAAAATGGACTAGTAATAAAGAG 1301
401 AspGIuPheArGmetProTyrLeuSerHisGIuInLeuProAlaGIyL 417
|||||
1302 GATGAGTTCAGGATGCTTATCTAAGCATCAGCAGCTGCTGTGGAAAT 1351
417 eleuPromeValProGIuValAlaInAlaValAGlyValSerGIuGIyH 434
|||||
1352 TCTTCCCATGTCGCCGAGGTCCGCCAGGCTGAGAGTTAGTCAAGGAC 1401
434 ishStrLysAspPheThrArGIAlaAlaProAsnProAlaLysAlaThr 450
|||||
1402 ATACACAAAGATTTTACACAGGACCTCCGAATCTGCCAAGGCCACG 1451
451 ValThrAlaMetIleAlaArgGIuLeuLeuTyrGIyGIyThrSerProTh 467
|||||
1452 GTACATGCCATGATAGCCCGAGAGTTGTGTATGGGGGCACCTGCCAC 1501
467 rAlaGIuThrIleuLysAsnAsnIleSerSerGIyHisValProHisG 484
|||||
1502 AGCCGAGACCATTTTAAAGATAACATCTCTTCAGGCCAGGTACCCCATG 1551
484 LyrProLeuThrArgProSerGIuGIuLeuAspTyrIleuSerArGIyGIu 500
|||||
1552 GACCTTCACAGAGACCTCTGACCACTGCACTATCTTCCAGAGTCCAG 1601
501 GIuPheGIuAlaGIuTyrLysAspPheProLysAsnAsnLysAsnGIuPh 517
|||||
1602 GGATTCAGAGTGTAGTAAAGACTTCCCAAAACAAACAGACGAAATP 1651
517 eValSerLeuIleAsnCySerSerGIuProProLeuIleSerHisGIyI 534
|||||
1652 TGATCTCTTATCATGTGCTCTCTCTGACCCACTCTGATCAGCATGTGTA 1701
534 legLysAspValGIuSerCyHisAspMetAlaAlaLeuAsnIleLeu 550
|||||
1702 TCGGCAAGGATGTGAGTCCCTGCCATGATATGGCTGCCGACATCTTA 1751
551 LysLeuLeuSerGIuLeuAspGIuInSerThrGIuMetProArGIyGIu 567
|||||
1752 AAGTTGCTGTGTAGTGGACCAACAAAGTACAGAGTGCACAAAGACAGG 1801
567 yAsnGIyProMetSerValCysGIyArGIyCys 577
|||||
1802 AAACGGACCAATGTCTGTGTGTGGAGGTGC 1832

```

seq\_name: /SID56/gcgdata/geneseq/geneseqn/NA2000.DAT:293327

seq\_documentation\_block:

ID 293327 standard; cDNA; 3260 BP.

AC 293327;

DT 04-JUL-2000 (first entry)

DE Human stauften cDNA.

XX Staufen: RNA binding protein; RNA/DNA hybrid; virus; retrovirus;  
 KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;  
 KW acquired immune deficiency syndrome; protease; ; human; ss.

OS Homo sapiens.

XX Key Location/Qualifiers  
 PH 409..1311  
 FT CDS /tag= a  
 FT /product= Staufen protein

FT

```

XX CA2238656-A1.
PN
XX
XX 22-NOV-1999.
PD
XX
XX 22-MAY-1998; 98CA-2238656.
PF
XX
XX 22-MAY-1998; 98CA-2238656.
PR
XX
XX (UYMO-) UNIV MONTREAL.
PA
XX Duchaine T, Wickham L, Cohen EA, Luo M, Desgroselliers L;
PI Moulund AJ;
PI
DR WPI: 2000-246924/22.
DR P-PSDB; Y83108.
XX
XX Isolated mammalian or Caenorhabditis elegans stauften proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)
PT
XX
XX Claim 4; Figure 1b; 96pp; English.
PS
XX
XX Staufen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC stauften with proteins with RNase or protease activity can be used to
CC treat virus infections. The RNase or protease activity of the fusion
CC protein prevents proper maturation of the virus. Mammalian stauften
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/MSF (human SP) and
CC MBP/MSF (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Stauften proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus. This
CC sequence is an alternatively spliced sequence to the one given in
CC 293322.
SQ
Sequence 3260 BP; 907 A; 795 C; 759 G; 799 T; 0 other;

```

alignment\_scores:

Quality: 2633.50 Length: 509

Ratio: 5.184 Gaps: 1

Percent Similarity: 99.804 Percent Identity: 99.804

alignment\_block:

US-09-316-048-6 x 293327 ..

Align seg 1/1 to: 293327 from: 1 to: 3260

```

69 G|uSerIleThrProThrValGIuLeuAsnAlaLeuCyMetLysLeuGI 85
|||||
370 GAAAGCATTAACCCCTACTGTAGAACTTAATGCACTGTGATGAAAGCTGG 419
85 yLysLysProMetTyrLysProValAspProTyrSerArgMetGIuInserT 102
|||||
420 AAAAACAACATGTATTAAGCCTGTGAACCTTACTCTCGGATGCAAGTCC. 468
102 hrTyrAsnTyrAsnMetArgGIyAlaTyrProProArGIyTyrPheTyr 118
|||||
469 .TATTACTCAACATGAGAGAGAGTGTCTTATCCCGAGAGTACTTTTAC 516
119 ProheProValProProLeuLeuTyrGIuValGIuLeuSerValGIyGI 135
|||||
517 CCATTTCAGTCCACCTTACTTATCAAGTGAAGTGAACCTTCTGTGGAGG 566
135 yGIuGIuPheAsnGIyLysGIyLysThrArgGIuAlaAlaLysHisAspA 152
|||||
567 ACAGCATTTTAATGCAAGAAAGACAAAGACAGAGCTGCGCAAAACAGATG 616

```

152 laAlaAlaLysAlaLeuArgIleLeuGlnAsnGluProLeuProGluArg 168  
 |||||  
 617 CTGCTGCCAAAGCGTTGAGAGATCTCTCACAATGAGCCCTGCCAGAGAG 666  
 169 LeuGluValAsnGlyArgGluSerGluGluAsnLeuAsnLysSerG1 185  
 |||||  
 667 CTGGAGGTGAATGAGAGAGAAATCCGAAAGAAATCTCAATAAATCTCA 716  
 185 uLLeSerGlnValPheGluIleAlaLeuLysArgAsnLeuProValAsp 202  
 |||||  
 717 AATAAGTCAAGTGTGTGAGATTGCACTTAACGGAACCTTGCCCTGGAAT 766  
 202 heGluValAlaArgGluSerGlyProProHISmetLysAsnPheValThr 218  
 |||||  
 767 TCGAGGTGGCCCGGAGAGTGGCCCAACCATGAAAGAACTTTGTGAGAC 816  
 219 LysValSerValGlyGluPheValGlyGluGlyGlyLysSerLysLy 235  
 |||||  
 817 AAGGTTTCGTTGGGGAGTTGTGGGGGAAGGTGAAGGGAAGAAAGCAAG 866  
 235 sLLeSerLysLysAsnAlaAlaIleAlaValLeuGluLeuLysLysL 252  
 |||||  
 867 GATTTCAAAGAAAAATGCCCATAGCTGTCTGTGAGGAGCTGAAGAGCT 916  
 252 euProLeuProAlaValAlaGluArgValLysProArgIleLysLys 268  
 |||||  
 917 TACCGCCCTGGCTGCAGTTGAACGAGTAAAGCTTGATCAAAAAAGAA 966  
 269 ThrLysProIleValLysProGlnThrSerProGluTyrGlyGlnGly 285  
 |||||  
 967 ACAAAAACCATAGTCAAGCCACAGAACCCCAATATATGCGCAGGAGAT 1016  
 285 eAsnProIleSerArgLeuAlaGlnIleGlnGlnAlaLysGluLysG 302  
 |||||  
 1017 CAATCCGATTACCCGACTGCGCCAGATCCAGACGCAAAAAAGAGAAAG 1066  
 302 LuProGluTyrThrLeuLeuThrGluArgGlyLeuProArgArgArgL 318  
 |||||  
 1067 ACCCAAGATACACGCTCTCTCACAGACGAGCCCTCCCGCCGACAGGAG 1116  
 319 PheValMetGlnValLysValGlyLysHisThrAlaGluGlyThrGly 335  
 |||||  
 1117 TTTGTGATCAGAGTGAAGCTTGAAACCACTGCGAAGAAAGGAGGAC 1166  
 335 rAsnLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleL 352  
 |||||  
 1167 CAACAAGAAAGGTGGCCAGCGCAATGCACCCGAGAACATGCTGAGATCC 1216  
 352 euGlyPheLysValProGlnArgGlnProThrLysProAlaLeuLys 368  
 |||||  
 1217 TTGGTTTCAAGTCCCGGACGCGGACCCACCAACCCGACCTGCAAGTCA 1266  
 369 GluGluLysThrProIleLysLysProGlyAspGlyArgLysValThr 385  
 |||||  
 1267 GAGGAGAAAGACCCCATTAAGAAACACGAGGATGGAAGAAAGTAACTT 1316  
 385 ePheGluProGlySerGlyAspGluAsnGlyThrSerAsnLysGluAsp 402  
 |||||  
 1317 TTTTGAACCTGGCTCGGGAGTGAATGGACTAGTAAAGAGGATG 1366  
 402 LuPheArgMetProTyrLeuSerHisGlnLeuProAlaGlyIleLeu 418  
 |||||  
 1367 AGTTCAAGATGCTTATCTTAAGTCATCAGACAGCTGCTCTGGAATCTT 1416  
 419 ProMetValProGluValAlaGlnAlaValGlyValSerGlnGlyHis 435  
 |||||  
 1417 CCCATGTGCTCCGAGGTCCGACGCTGTAAGAGTTAGTCAAGAGACATCA 1466  
 435 sThrLysAspPheThrArgAlaAlaProAsnProAlaLysAlaThrVal 452  
 |||||  
 1467 CACCAAGATTATTACAGGCGAGCTCCGAAATCTCTCCAAAGGCCAGCTTA 1516

452 hrAlaMetIleAlaArgGluLeuLeuTyrGlyGlyThrSerProThrAla 468  
 |||||  
 1517 CTGCCATGATACCCGAGAGTGTGTGTATGGGGGACACCTCGGCCACAGCC 1566  
 469 GluThrIleLeuLysAsnAsnIleSerSerGlyHisValProHisGlyPr 485  
 |||||  
 1567 GAGACCATTTTAAAGAAATTAACATCTTTCAGGCCACGTAACCCATGAGCC 1616  
 485 oLeuThrArgProSerGluGlnLeuAspTyrLeuSerArgValGlnGlyP 502  
 |||||  
 1617 TCTCAGAGAACCTCTGAGCACTGAGCTATCTTTCAGAGTCCAGGAGAT 1666  
 502 heGlnValGluTyrLysAspPheProLysAsnAsnLysAsnGluPheVal 518  
 |||||  
 1667 TCCAGGTGAATACAAAGACTTCCCAAAAAACAAAGAAAGAAATTTGTA 1716  
 519 SerLeuIleAsnCysSerSerGlnProProLeuIleSerHisGlyIleG1 535  
 |||||  
 1717 TCTCTTATCAATTGCTCTCTCAGCCACCTCTGATCAGCCATGATTCGG 1766  
 535 yLysAspValGluSerCysHisAspMetAlaAlaLeuAsnIleLeuLysL 552  
 |||||  
 1767 CAAGATGTGAGTCTCTGCCATGATATGCTGCTGCAACATCTTAACT 1816  
 552 euLeuSerGluLeuAspGlnGlnSerThrGluMetProArgThrGlyAsn 568  
 |||||  
 1817 TGCTGTCTGAGTTGGACCAACAAAGTACAGAGATGCCAAGAACAGGAAC 1866  
 569 GlyProMetSerValCysGlyArgCys 577  
 |||||  
 1867 GGACCAATGTCTGTGTGGAGGTGC 1893

seq\_name: /SIDS6/gcgdata/geneseq/geneseqn/NA2000.DAT:293328

seq\_documentation\_block:

ID Z93328 standard; cDNA; 2859 BP.

AC Z93328;

DT 04-JUL-2000 (first entry)

DE Mouse staufen cDNA.

KW Staufen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;

KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;

KW acquired immune deficiency syndrome; protease; mouse; ss.

OS Mus musculus.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

CA2238656-A1.  
 22-NOV-1999.  
 22-MAY-1998; 98CA-2238656.  
 22-MAY-1998; 98CA-2238656.  
 (UYMO-) UNIV MONTREAL.  
 Duchaine T, Wickham L, Cohen EA, Luo M, Desgroselliers L;  
 Moulard AJ;  
 WPI: 2000-246924/22.  
 P-PSDB: Y83024.  
 Isolated mammalian or Caenorhabditis elegans staufen proteins useful  
 for treating retroviral infections especially human immunodeficiency  
 virus infections (HIV)

XX Claim 4; Figure 1c; 96pp; English.

PS Staufen is a RNA binding protein which interacts with double stranded  
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA  
CC transport and localization. Mammalian SP contain multiple ds  
CC RNA-binding domains. Chimeric proteins comprising portions of  
CC Staufen with proteins with RNase or protease activity can be used to  
CC treat virus infections. The RNase or protease activity of the fusion  
CC protein prevents proper maturation of the virus. Mammalian staufen  
CC proteins seem to recognise double stranded RNA structure rather than  
CC any sequence specific position. 2 bacterially-expressed fusion  
CC proteins used in an RNA-binding assay (his/hsp (human SP) and  
CC MBP/msp (murine SP)) both proteins strongly bound double stranded  
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and  
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the  
CC treatment of RNA virus infections, especially those caused by  
CC retroviruses, in particular human immunodeficiency virus.

XX Sequence 2859 BP; 747 A; 750 C; 712 G; 650 T; 0 other;

alignment\_scores:                      Quality: 2422.50                      Length: 552  
   Ratio: 4.750    Gaps: 2  
Percent Similarity: 92.391              Percent Identity: 85.145

alignment\_block:  
US-09-316-048-6 x 293328 ..

Align seg 1/1 to: 293328 from: 1 to: 2859

```
26 SerLeuSerGlnProLeuMetSerIleProSerThrThrSerSerIle 42
157 TCCCTCCTCGGTCGGCCCTTTTCCTCGCGCTTCCTCCTCCTTC 206
42 uProSerGluAsnAlaGlyArgProIleGlnAsnSerAlaLeuProSerA 59
207 ACCTCCTCGCGCGCCCAAGA.....CCGCGG 235
59 IAserIleThrSerThrSerAlaAlaGlnSerIleThrProThrVal 75
236 CCCCAGGAGACGCTCTGGGAAACACACCAAGATATACCTTACCAT 285
76 GluLeuAsnAlaLeuGlyMetLeuGlyLysLysProMetIleArg 92
286 GAGCTCAATGCACTGTGTGAAGTGAAGAAACCAATGATTAAGCC 335
92 oValAspProTyrSerArgMetGlnSerThrTyrAsnTyrAsnMetArg 109
336 CGTGACCCCACTCTCGATGCAATGCACCTACAGTATGCGATGCGTG 385
109 LysIleAlaTyrProProArgTyrPheTyrProDheProValProProLeu 125
386 GAGTGCTATCCGCCAGATCTTACCATTTCCAGTCCACACTT 435
126 LeuTyrGlnValGluLeuSerValGlyGlnGlnPheAsnGlyLysG 142
436 CTCTACCAAGTGAAGCTCTCGTGGGGGACACAGTTTAATGGAAAG 485
142 YLysThrArgGlnAlaAlaLysHisAspAlaAlaLysAlaLeuArgI 159
486 AAGATGAGAGCCACCGCTGAACACGATGCCCTCGCGCTGAGGA 535
159 ILeuGlnAsnGluProLeuProGluArgLeuGluValAsnGlyArgGlu 175
536 CTCCTGAGAGTGAACCCCTGCCAGAAAGTTGAGAGTAATGGAAGAA 585
176 SerGlnGlnGluAsnLeuAsnLysSerGluIleSerGlnValPheGlu 192
586 GCAGAGGAAAGAAACCTCAATTAATGCAATTAAGCAAGTGTGTAAT 635
192 eAlaLeuLysArgAsnLeuProValAsnPheGlnValAlaArgLys 209
```

```
636 TCGCTGAACCGCAATTTGCTGTGATTTTGAGGTGGCCGGAGAGTG 685
209 LysProProHISMetLysAsnPheValThrLysValSerValGlyGluPhe 225
686 GCCCACCACCAATGAAGAACTTGTGACAGGGGTTTGTAGTTGGGAATTT 735
226 ValGlyGluLysGlnGlyLysSerLysLysIleSerLysLysAsnAlaI 242
736 GTAGGGGAAGAGAGGAAAGCAAGCAAGATCTCCAGAAAGATGGCGC 765
242 AlAlaValLeuGlnGluLeuLysLysLysLeuProProLeuProAlaVal 259
786 CAGGCGCTGTCTGAGCAGAGCTTAGAGAGCTGCACCCCTCGCTGTGG 835
259 LysArgValLysProArgIleLysLysLysThrLysProIleValLysPro 275
836 AGCGAGTGAAGCCCAATCAAGAAAGTCAAGCCCTCCCAAG... 882
276 GluThrSerProGluTyrGlyGlnGlyIleAsnProIleSerArgLeuAl 292
883 ...ACAGCCCCGATTTAGGCCAAGGATGAATCTTATGTAGACTTGC 929
292 agnIleGlnGlnAlaLysLysGluLysGluProGluTyrThrLeuLeuT 309
930 ACAGATCCACAGCAGCAAAAAAGAGAAAGAGCCAGATACATGCTCTTA 979
309 hrcLuarGlyLysProArgArgArgGluPheValMetGlnValLysVal 325
980 CAGAACAGGCTCTTCCACGTCCAGAGGAGTGTGATGACAGTAAGGTT 1029
326 GlyAsnHisThrAlaGluGluGlyThrGlyThrAsnLysLysValAlaLys 342
1030 GGGCATCACCTGACAGAGAGAGGTGGTACCAATTAAGAGGTGGCCAGCG 1079
342 gAsnAlaAlaGluAsnMetLeuGluLysLeuGlyPheLysValProGlnA 359
1080 TAATGCTGTGAGAACATGCTGAGATCTGGGGTTCAAAAGTCCCAAG 1129
359 rglInProThrLysProAlaLeuLysSerGluLysThrProIleLys 375
1130 CGCAGCTGCCAAGCCAGCACTCAATCAAGAGAAAGACCTCCAGTAAAG 1179
376 LysProGluAspGlyArgLysValThrPhePheGluProGlySerGlyAs 392
1180 AAACCGAGAGACGAGAGAAAGTACCTTTTGAACCTTACCTGGGGA 1229
392 pGluAsnGlyThrSerAsnLysGluAspGluPheArgMetProTyrLeus 409
1230 TGAATAATGGAAGTAAAGTAAAGAGAGAGTTCAGAGTCCCTATCTTA 1279
409 erHisGlnGlnLeuProAlaGlyIleLeuProMetValProGluValAla 425
1280 GCCATCAGCAGCTGCCAGTGTGAATTCCTCCCATGTGCGGGAAGTTGCC 1329
426 GlnAlaValGlyValSerGlnGlnLysHisThrLysAspPheThrArg 442
1330 CAGGCTGTGGGGTGTAGTCAAGACACACACCAAAATTTACACAGGCG 1379
442 aAlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArgLul 459
1380 AGCTCCAAATCTGCCAAGCAAGCAAGGTAACTGCCATGATGCGCAAGT 1429
459 euleuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysAsnAsn 475
1430 TGTGTGTGCGGGGACCTCCACACAGCCAGACCAATTTAAAGAGTAAC 1479
476 IleSerSerGlyHisValProHisGlyProLeuThrArgProSerGlu 492
1480 ATCTCTTCAGGCAAGTACCCATGAGCTCGCACTTAGACCTCTGAGCA 1529
492 nLeuAspTyrLeuSerArgValGlnGlyPheGlnValGluTyrLysAsp 509
```

```

1530 ACTGACTACCTTTCCAGAGCCAGGATTCAGTTGAATACAAAGATT 1579
509 hnpProLysAsnAluValSerLeuIleAsnYsSer 525
|||||
1580 TTCACAGAACAAACAGACGATGTATCTCTCATCACTGCTCTCA 1629
526 GluProLeuIleSerHisGlyIleGlyLysAspValGluSerCysH 542
|||||
1630 CAGCCGCTCTGCTGATCATGCGCAAGATGTGAGTCTGTCA 1679
542 sAspMetAlaIleAluAsnIleLeuLysLeuSerGluLeuAspGlnG 559
|||||
1680 TGAATATGCTGACATGAACATTTTAAAGCTGCTGTGAGTTGGACCA 1729
559 lSerThrGluMetProArgThrGlyAsnGlyPrometSerValCysGly 575
|||||
1730 AGAGCACACAGATGCCAAGACAGAAATGGACAGTTTCAGCGTGGG 1779
576 ArgCys 577
|||||
1780 AGGTGC 1785

seq_name: /SIBS6/gcdata/geneseq/geneseqn/NA1999.DAT:X39991
seq_documentation_block:
ID X39991 standard; DNA: 1127 BP.
AC X39991;
XX
XX 02-JUL-1999 (first entry)
XX
XX Prostate cancer associated gene.
XX
XX Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KM prostate cancer; ss.
XX
XX Homo sapiens.
XX
XX W09904265-A2.
XX
XX 28-JAN-1999.
XX
XX 15-JUL-1998; 98WO-US14679.
XX
XX 22-JUN-1998; 98US-0102322.
XX
XX 17-JUL-1997; 97US-0886164.
XX
XX 10-OCT-1997; 97US-0061599.
XX
XX 10-OCT-1997; 97US-0061765.
XX
XX 10-OCT-1997; 97US-0948705.
XX
XX 11-OCT-1997; 97GB-0021697.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
XX
XX Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ,
PI Pfeundschn M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
XX WPI: 1999-132448/11.
XX
XX New isolated cancer associated nucleic acids and polypeptides -
XX isolated using sera from cancer patients, used to develop products
XX for the diagnosis, monitoring or treatment of cancers
XX
XX Claim 67: Page 625-626; 787pp; English.
XX
XX The invention relates to a method for diagnosing a disorder characterised
XX by expression of a human cancer associated antigen precursor coded for by
XX a nucleic acid molecule (NAM). The method comprises: (a) contacting a
XX biological sample isolated from a subject with an agent that specifically
XX binds to the NAM, an expression product or a fragment of an expression
XX product complexed with an HLA molecule; and (b) determining the
XX interaction between the agent and the NAM or the expression product as a

```

```

CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
XX Sequence 1127 BP; 349 A; 265 C; 307 G; 203 T; 3 other;
50

alignment_scores:
Quality: 1880.00 Length: 375
Ratio: 5.054 Gaps: 1
Percent Similarity: 99.200 Percent Identity: 98.400

alignment_block:
US-09-316-048-6 x X39991 ..

Align seg 1/1 to: X39991 from: 1 to: 1127

135 GYGINGInPheAsnGlySerGlyLysThrArgGlnAlaAlaHisAs 151
|||||
3 GGACAGCAATTTAATGGCAAGGAAAGACAGACAGGCTGCGAAACAGCA 52
151 PALAAlaAlaLysAlaLeuArgIleLeuGlnAsnGluProLeuProGluA 168
|||||
53 TGCTGCTGCCAAGCCTTGAGGATCTCTGAGAAATAGGCCCTGCCAGAA 102
168 rGluGluValAsnGlyArgGluSerGluGluGluAsnLeuAsnYsSer 184
|||||
103 GGCTGGAGGTGAATGGAAGAAATCCGAAAGAAAGAAATCTCAATAATCT 152
185 GluIleSerGlnValPheGluIleAlaLeuLysArgAsnLeuProValAs 201
|||||
153 GAATAGTCGAAGTGTGAGATGCACTTAACGGAACCTGCTGTGA 202
201 nPheGluValAlaArgGluSerGlyProProHisMetLysAsnPheValT 218
|||||
203 TTTGAGGTGGCCCGGAGAGATGGCCACCACCAATGAAGAACTTTGTGA 252
218 hLysValSerValGlyGluPheValGlyGluGlyLysSerLys 234
|||||
253 CCAAGGTTTCGGTTGGGAGTTGTGGGGAAGTGAAGCAAAACCAAG 302
235 LysIleSerLysAsnAlaAlaIleAlaValLeuGluGluLeuLysL 251
|||||
303 AAGATTTCAAAGAAATAATCCGCATAGCTGTCTTGAGGAGCTGAAGAA 352
251 sLeuProProLeuProAlaValAlaGluArgValLysProArgIleLysL 268
|||||
353 GTTACCGCCCTGCTGCGAGTTGACGAGTAAGCCTAATAATCAAAAAA 402
268 ySThrLysProIleValLysProGlnThrSerProGluThrGlyGlnGly 284
|||||
403 AAACAACCAACCATAGTCACAGCCACAGACAAAGCCCAAAATATGGCCG 452
285 lLeasnProIleSerArgLeuAlaGlnIleGlnGlnAlaLysLysGly 301
|||||
453 ATCAATCCGATTAGCCGACTGCGCCAGATCCAGACGCAAAAAAGAGAA 502
301 sGluProGluThrThrLeuLeuThrGluArgGlyLeuProArgArgArg 318
|||||
503 GGAGCCAGAGTACAGCGCTCTCCACAGAGGAGGCTCCCGCGCGAGGG 552
318 lnPheValMetGlnValLysValGlyAsnHisThrAlaGluLysThrGly 334
|||||
553 AGTTTGATGTCAGGTGAGGAGTTGGAAACACACATGTCACAAAGAACGG 602
335 ThrAsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGlu 351
|||||
603 ACCAACAGAAAGGTGGCCAAAGCGCAATGACGCGAGAAACATGCTGAGAT 652

```

```

351 eLeuGlyPheLysValProGlnArgGlnProThrLysProAlaLeuLysS 368
    |||||||
653 CTTGGTTTCAAGTCCCGCAGCGCAGCCACCAAAACCGCAGTCAAGT 702
    |||||||
368 eGluGluLysThrProIleLysLysProGlyAspGlyArgLysValThr 384
    |||||||
703 CAGAGGAGAAGACACCCATMAAGAAACAGGGGATGMAAGAAAGTAAAC 752
    |||||||
385 PhePheGluProGlySerGlyAspGlyAsnGlyThrSerLysGluAs 401
    |||||||
753 TTTTGTGAACCTGGCTCTGGGATGAAATGGAGCTAGTAATATAAGGGA 802
    |||||||
401 pGluPheArgMetProTyrLeuSerHisGlnGlnLeuProAlaGlyIle 418
    |||||||
803 TGGGTTCAAGATGCTTATCTATAGTCATCAGCAGCTCCCGCTGGAAATTC 852
    |||||||
418 euProMetValProGluValAlaGlnAlaValAlaGlyValSerGlnGlyHis 434
    |||||||
853 TTCCATGTGTGCCCCGAGGTCGCCAGGCTGTAGAGTTAGTCAAGGACAT 902
    |||||||
435 HisThrLysAspPheThrArgAlaAlaProAsnProAlaLysAlaThrVa 451
    |||||||
903 CACACCAAGATTTTACAGGAGCAGCTCCGAATCTGCCAAGGCCACGGT 952
    |||||||
451 lPheAlaMetIleAlaArgGluLeuLeuTyrGlyGlyThrSerProThra 468
    |||||||
953 AACTGGCATATAGCCCGAGAGTGTGTGTGGGGGACACTCCGCCACAG 1002
    |||||||
468 lA.GluThrIleLeuLysAsnAsnIleSerSerGlyHisValProHisG1 484
    |||||||
1003 CCCGAGACCATTTTAAAGATAACATCTCTTCAGGCCACGACCCCATGG 1052
    |||||||
484 yProLeuThrArgProSerGlnGlnLeuAspTyrLeuSerArgValGlnG 501
    |||||||
1053 ACCTCTACAGACCCCTMTGAGCAACTGGACTATCTTTNCAGAGTCCAGG 1102
    |||||||
501 lYPheGlnValGluTyrLysAsp 508
    |||||||
1103 GATTNCAGGTTGAATACCAAGAC 1125
    |||||||

seq_name: /SID56/gcgcdata/geneseq/geneseqn/NA1999.DAT: X39992
seq_documentation_block:
ID      X39992 standard; DNA: 773 BP.
AC      X39992:
XX
XX
DT      02-JUL-1999 (first entry)
XX
XX
DE      Prostate cancer associated gene.
XX
XX
KM      Cancer associated antigen; diagnosis; research; treatment; human;
KM      breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KM      prostate cancer; ss.
XX
XX
OS      Homo sapiens.
XX
XX
PN      W09904265-A2.
XX
XX
PD      28-JAN-1999.
XX
XX
PF      15-JUL-1998; 98MO-US14679.
XX
XX
PR      22-JUN-1998; 98US-0102322.
PR      17-JUL-1997; 97US-0896164.
PR      10-OCT-1997; 97US-0061599.
PR      10-OCT-1997; 97US-0061765.
PR      10-OCT-1997; 97US-0948705.
PR      11-OCT-1997; 97GB-0021697.
XX
XX
PA      (LUDW-) LUDWIG INST CANCER RES.
XX
PI      Chen Y, Gout I, Gure A, O'Hare M, Odata Y, Old L;

```

```

PI      Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI      Tureci O;
XX
XX
DR      WPI; 1999-132448/11.
XX
XX
PT      New isolated cancer associated nucleic acids and polypeptides -
PT      isolated using sera from cancer patients, used to develop products
PT      for the diagnosis, monitoring or treatment of cancers
XX
XX
PS      Claim 67; Page 626; 787pp; English.
XX
XX
CC      The invention relates to a method for diagnosing a disorder characterised
CC      by expression of a human cancer associated antigen precursor coded for by
CC      a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC      biological sample isolated from a subject with an agent that specifically
CC      binds to the NAM, an expression product or a fragment of an expression
CC      product complexed with an HLA molecule; and (b) determining the
CC      interaction between the agent and the NAM or the expression product as a
CC      determination of the disorder. The products and methods can be used in
CC      the diagnosis, monitoring, research, or treatment of conditions
CC      characterised by the expression of various cancer associated antigens.
CC      The invention provides nucleic acid sequences and encoded polypeptides
CC      which are cancer associated antigen precursors expressed in human breast
CC      cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC      lung cancer.
XX
XX
SQ      Sequence 773 BP; 233 A; 204 C; 199 G; 134 T; 3 other;

alignment_scores:
    Quality: 1292.00      Length: 257
    Ratio: 5.087          Gaps: 1
    Percent Similarity: 98.833      Percent Identity: 97.665

alignment_block:
US-09-316-048-6 x X39992 ..

Align seg 1/1 to: X39992 from: 1 to: 773

253 ProProLeuProAlaValGluArgValLysProArgIleLysLysLysTh 269
    |||||||
3   CCGCCCCCTGCCTGCAGTTGAACGAGTAAGCTGAAATCAAAAAGAAAC 52
    |||||||
269 rLysProIleValLysProGlnThrSerProGluTyrGlyGlnGlyIleA 286
    |||||||
53 AAAACCCATGATGTCAGCCACAGACAGCCCAAGATATAGCGAGGATCA 102
    |||||||
286 snProIleSerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGlu 302
    |||||||
103 ATCGATTAGCGGACGCTGCGCAGATCCAGCAGCGCAAAAAGGAGAGAG 152
    |||||||
303 ProGluTyrThrLeuLeuThrGluArgGlyLeuProArgArgArgGluPh 319
    |||||||
153 CCAAGATACACGCTCTCCACAGACGAGGCTCCCGCGCCGACAGGAGTT 202
    |||||||
319 eValMetGlnValLysValGlyAsnHisThrAlaGluGlyThrGlyThra 336
    |||||||
203 TGTGATCAGGTGAAGTTGGAACCACTGCAAGAGGAGGAGCA 252
    |||||||
336 snLysLysValAlaLysArgAsnAlaLacLysAsnMetLeuGluIleLeu 352
    |||||||
253 ACAAGAGAGGTGGCCAGCGCATGTCAGCCGAGAACATGCTGGAATCCTT 302
    |||||||
353 GlyPheLysValProGlnArgGlnProThrLysProAlaLeuLysSerG1 369
    |||||||
303 GGTTCAAAGTCCCGCAGGCGCACCCCAACCCACACTCAAGTCAAGA 352
    |||||||
369 uGluLysThrProIleLysLysProGlyAspGlyArgLysValThrPheP 386
    |||||||
353 GGAGAGACACCCATTAAGAACACCGGGGATGGAAGAAAGTAACTTTT 402
    |||||||
386 heGluProGlySerGlyAspGlyAsnGlyThrSerAsnLysGluAspGlu 402
    |||||||

```

403 TTGACCTGGCTCTGGGATGTAAATGGACCTACGTAATTAAGACGATGAG 4522  
440 PheargmetProtyrleuSerHisGlnGlnLeuProAlaGlyIleLeuPr 4199  
443 TTGAGGATGGCTTATCTTAAGTATCATGACGACGTGCTGTGGAAATCTTCC 5020  
419 omcValProGluValAlaGlnIleValGlyValSerGlnIyHisIst 4366  
503 CATGGTGCCCGAAGTCGCCCAAGCTGTAGGAGTTATGTCAGGACATACA 5522  
436 hrllysasPheThrArgAlaAlaIaProAsnProAlaIyAlaThrValThr 4523  
553 CCAAGATTTTACAGAGGACGCTCGCAATCTGCCCAAGGCCACGTAAC 6022  
453 AlamelIleAlaArgGluLeuLeuTyrGlyIyThrSerProThrAla.G 4699  
603 GCCATGATAGGCCGGAAGTGTGTTGATGGGGGACACCTCGCCACAGCCCG 6522  
469 luvhrIleuLeuLysAsnAsnIleSerSerGlyHisValProHisGlyPro 4855  
653 AGACCAATTTAAAGATTAACATCTCTTTAGGCCACAGTACCCCATGACCT 7022  
486 leuThrArgProSerGlnGlnLeuAspTyrIleuSerArgValGlnGlyph 5020  
703 CTCACAGGAGACCCCTMTGACCACTGACATATCTTTNCAAGAGTCGAGGATT 7522  
502 eglInvalGluTyrLysAsp 508  
153 NCAGGTGTGAATACCAAGAC 771

```
seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1999.DAT:X400000
seq_documentation_block:
```

XX	XA0000;	
AC		
XX		
DT	02-JUL-1999	(first entry)
XX		
DE	Prostate cancer associated gene.	
XX		
KW	Cancer associated antigen; diagnosis; research; treatment; human; breast cancer; colon cancer; gastric cancer; lung cancer; prostate cancer; ss.	
XX		
OS	Homo sapiens.	
PN	WO9904265-A2.	
XX		
PD	28-JAN-1999.	
XX		
PF	15-JUL-1998;	98WO-US14679.
XX		
PR	22-JUN-1998;	98US-0102322.
PR	17-JUL-1997;	97US-0896164.
PR	10-OCT-1997;	97US-0061599.
PR	10-OCT-1997;	97US-0061765.
PR	10-OCT-1997;	97US-0948705.
PR	11-OCT-1997;	97GB-0021697.
XX		
PA	(LUDMW-) LUDMWIG INST CANCER RES.	
XX		
PI	Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ; Pfeundschuh M, Sahin U, Scanlan MJ, Stockert E; Tureci O;	
PI		
XX		
DR	WPI: 1999-132448/11.	
XX		
PT	New isolated cancer associated nucleic acids and polypeptides - isolated using sera from cancer patients, used to develop products for the diagnosis, monitoring or treatment of cancers	
PT		
XX		
PS	Claim 67; Page 629; 787pp; English.	

XX The invention relates to a method for diagnosing a disorder characterised  
CC by expression of a human cancer associated antigen precursor coded for by  
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
CC biological sample isolated from a subject with an agent that specifically  
CC binds to the NAM, an expression product or a fragment of an expression  
CC product complexed with an HLA molecule; and (b) determining the  
CC interaction between the agent and the NAM or the expression product as a  
CC determination of the disorder. The products and methods can be used in  
CC the diagnosis, monitoring, research, or treatment of conditions  
CC characterised by the expression of various cancer associated antigens.  
CC The invention provides nucleic acid sequences and encoded polypeptides  
CC which are cancer associated antigen precursors expressed in human breast  
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
CC lung cancer.

alignment\_scores:

Quality:	1257.50	Length:	281
Ratio:	4.800	Gaps:	2
Percent Similarity:	93.238	Percent Identity:	90.391

alignment\_block:

US-09-316-048-6 X X400000

Align seg 1/1 to: X40000 from: 1 to: 844

189 ValPheGluIleAlaLeuLysArgAsnLeuProValAsnPheGluValAl 205  
 |||||  
 3 GTGTTTGAGATTGCACCTTAACGGAACCTTGCTGCAATTTCGAGGTGCC 52

205 aarggluserglyproprohismetlysasnphavalthrlysvalsev 222

53 CCGGAGAGTGGCCACCCACATGAGAACTTGTGACCAAGTTTCGG 102

222 a1G1G1uPhevalG1G1uG1G1uG1G1uLysSerLysLysIleSerLys 238

103 TTGGGGAGTTTCTGCGGGGAAGGTGAAGGGAAGCAAGAGATTTCAAAG 1522

239 LYSASNALALALILEALVALLEUGLUGLUGLULEULYSLSLEUPROPROLE 255

153 AAAAATGCCGCCATAGCTGTTCTTGAGGAGCTGAGAGAGTTACGCCCTT 2022

255 uproAlaValGluArqValLvsProArqIleLvsLvsLvsThrlvsProI 272

203 GCCTGCAGTTGAACGAGTAAAGCCTAGATCAAAAAGAAACAAACCA 252

277 |eVa|Iv\$ProG|ntHr\$erProG|uTvrG|vG|ng|vT|eAsnProT|e 288

[illegible]

305  
306

203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000

2000

[illegible]

1000

[illegible][illegible]

403 G1BGCCHAGCUCCHAI1BCACGCGHGHCH1GCIUBGMHLCI1UBOI1ICGN 202

355 svaLProGlnArgGlnProGlnIrrLysProAlaLeuLysSerGluGluLysr 3/2

503 AGTCCCGCAGGCGAGCCCAACCCGACATCAGTTCAGAGGAGA 552



```

372 hrProIleLysLysProGlyAspGlyArgLysValThrPhePheGluPro 388
|||||
553 CACCATTAAAGAAACACGAGGATGGAAGAAAGTAACCTTTTGAACCT 602
|||||
389 GLYSerGlyAspGluAsnGlyThrSerAsnLysGluAspGluPheArg 405
::: |||||||
603 GCCTCTGGGGATGATAAATGGAGTAAATTAAGAGATGAGTTCAGAT 652
|||||
405 tProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuProMetValP 422
|||||
653 GCCTTATCTAGTCAATCAGCAGCTGCTGCTGAATCTTCCATGCTGCT 702
|||||
422 roGluValAlaGlnAlaValGly.ValSerGlnGlyHisThrLysAs 438
|| |||||||
703 CCGANGTGGCCCAAGCTGTAGGAAGTATGACAGACATNACACCCAGA 752
|||||
438 pPheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMet 455
|||||
753 TTTTACCGACGACGCTT...CGAATCTTGGCCAGCNCNGTACTGCTGA 799
|||||
455 leAlaArgGluLeuLeuTyrGlyGlyThrSerProThrAla 468
|||||
800 TACCCCAAGTGTGTGATTTGGGGCAGCTTGGCCCAAGGCC 840
|||||
seq_name: /SID6/gcgsdata/geneseq/geneseqn/NA1999.DAT.X40002
seq_documentation_block:
ID X40002 standard; DNA; 946 BP.
XX
AC X40002;
XX
DT 02-JUL-1999 (first entry)
XX
DE Prostate cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9904265-A2.
PD
PD 28-JAN-1999.
PD
PE 15-JUL-1998; 98WO-US14679.
XX
XX 22-JUN-1998; 98US-0103322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfeundschnh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI: 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Claim 67: Page 629-630; 787bp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression

```

```

CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
SQ Sequence 946 BP; 292 A; 208 C; 259 G; 170 T; 17 other:
XX
alignment_scores:
Quality: 1255.50 Length: 312
Ratio: 4.516 Gaps: 3
Percent Similarity: 89.103 Percent Identity: 85.577
alignment_block:
US-09-316-048-6 x X40002 ..
Align seg 1/1 to: X40002 from: 1 to: 946
135 GlyGlnGlnPheAsnGlyLysGlyLysThrArgGlnAlaLysHisAs 151
|||||
3 GGACAGCAATTTAATGCAAAAGAAAGAACAGACAGCTCGCAAAACAGA 52
|||||
151 pAlaAlaAlaLysAlaLeuArgIleLeuGlnAsnGluProLeuProGlu 168
|||||
53 TGCTGTGCCAAAGCGTTGAGAGATCTGCAAGATGAGCCCTGCGAAGA 102
|||||
168 rGluGluValAlaAsnGlyArgGluSerGluGluGluAsnLeuAsnLysSer 184
|||||
103 GGCTGGAGGTAAATGCAAGACATCCGAGAGAAATAATCTAATTAATCT 152
|||||
185 GluIleSerGlnValPheGluIleAlaLeuLysArgAsnLeuProValAs 201
|||||
153 GAATTAAGTCAAGTGTGAGATGCACTTAACGAACTTGCCGTGGA 202
|||||
201 nPheGluValAlaArgGluSerGlyProProHisMetLysAsnPheValT 218
|||||
203 TTTGAGGTGGCCCGGAGAGTGGCCACCCACATGAAAGAACTTTGGA 252
|||||
218 hrLysValSerValGlyGluPheValGlyGluGlyLysSerLys 234
|||||
253 CCAAGGTTTCGTTGGGAGTGTGGGGGAAGTGAAAGGAAAGCAAG 302
|||||
235 LysIleSerLysAsnAlaAlaIleAlaValLeuGluLeuLysLys 251
|||||
303 AAGATTTCAAGAAATAATGCCCATAGCTGTCTTGAGAGAGTGAAGAA 352
|||||
251 sLeuProProLeuProAlaValGluArgValLysProArgIleLysLysL 268
|||||
353 GTTACCGCCCTGCTGCAAGTGAACGAGTAACCTGAATCAAAAAGA 402
|||||
268 yThrLysProIleValLysProGlnThrSerProGluTyrGlyGlnGly 284
|||||
403 AAACAAAACCATAGTCAAGCCACAGACACCAAGAAATATGCGCAGGG 452
|||||
285 IleAsnProIleSerArgLeuAlaGlnIleGlnGlnAlaLysLysGly 301
|||||
453 ATCAATCCGATTAGCCGACGCGCCAGATCCAGCGCAAAAAGAGAA 502
|||||
301 sGluProGluTyrThrLeuLeuThrGluArgGlyLeuProArgArgArg 318
|||||
503 GGAGCCAGAGTACACGCTCTCACAGAGCGAGGCTTNCGGCGCGCAGG 552
|||||
318 lPheValMetGlnValLysValGlyAsnHisThrAlaGluGlyThrG1 334
|||||
553 AGTTTGATGATCAGGTGAAGTTGGAACCACTTGCAGAAAGAACGG 602
|||||
334 yThrAsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGlu 351
|||||

```

```

603 CACCAAGAGAGTGGCCAGCCAGCATGACCAGCAACATGCTGTGAGA 652
351 leuenuLyhe LyValProGlnArgGlnProThrIysProAlaLeuY 367
653 TCCTTGCTTCAAAAGTCCCGCAGCCAGCCAGCAACCCGAGCTTAA 702
367 sSerGluGluYThrProIleLysIysProGlyAspGlyArgLyValT 384
703 GTCAGAGNAGAGACCCCATTAAGAAACCAACGATGAGAGAAAGTAN 752
384 hrPheheGluProGlySerGlyAspGluasnGlyThrSeranlySglu 400
753 CMTTGTGAAGCTGGCTNTTGGGA...TTAAATGGCTGTANTTAAGAG 799
401 AspGluPheArgMetProTyrIleuSerHisGlnGlnLeuProAlaGly 417
800 GATGAGTTCAGATGACGTCATTAAGTCAATACACTTCTCT...GCTGGA 848
417 leuPProMetValProGluValAlaGlnAlaValGlyValSerGlnGly 434
849 TCCTTCCATGGGGCGGAGGTCGCCAGCTTTAGAGATTATCANGCCN 898
434 lshIsthrLysAspPheThrArgAlaAlaPro 444
899 TTCNCACCAAGAAATTTCCCGCGCAGTTTCCA 930
seq_name: /SID6/gcdata/geneseq/geneseqn/NA2000.DAT:279981

```

```

seq_documentation_block:
ID 279981 standard; cDNA; 660 BP.
XX
AC 279981;
XX
DT 07-APR-2000 (first entry)
XX
DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:65.
XX
KW Human; gene expression product; diagnosis; tumour; colon cancer;
KW colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW cytosolic; sarcoma; breast cancer; neoplasia; dysplasia;
KW hyperplasia; ds.
XX
OS Homo sapiens.
XX
PN M09964576-A2.
XX
PD 16-DEC-1999.
XX
PF 09-JUN-1999; 99WO-IB01062.
XX
PR 10-JUN-1998; 98US-0088801.
XX
PA (FARB ) BAYER CORP.
XX
PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catino TJ, Dertl A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;
XX
WPI: 2000-087220/07.
XX
DR Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer
XX
PS Claim 1; Page 158; 469pp; English.
XX
CC 279917 to 280766 represent double stranded cDNA clones isolated from the
CC human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antitense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g. carcinomas
CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used

```

CC for developing agents for the diagnosis and treatment of disorders  
 CC involving unwanted cell proliferation, such as neoplasia, dysplasia or  
 CC hyperplasia.  
 CC  
 SQ Sequence 660 BP; 133 A; 172 C; 159 G; 174 T; 22 other:

alignment\_scores:  
 Quality: 697.00 Length: 204  
 Ratio: 4.052 Gaps: 3  
 Percent Similarity: 84.314 Percent Identity: 75.490

alignment\_block:  
 US-09-316-048-6 x 279981/rev

Align seg 1/1 to reverse of: 279981 from: 1 to: 660

```

279 ProGUtYrGlyGlnGlyLeuAsnProIleSerArgLeuAlaGln...11 294
607 CCCCANGTTGTCGGAAGTTGATTCCTCCTTACAGGCGCCAGTGNGT 558
294 eGlnGlnAlaLysLysGluLysGluProGluTyrThrLeuLeuThrGluA 311
557 GGANTTCGCCNTAAGCGTGTCCGCGCAAGGTACACGTTCTTCANGAAG 508
311 rGlyLeuProArGArGArGlu.PheValMetGlnValLysValGlyLys 327
507 CNAAGCTTCGCGNGCGNAGAGAGTTGTATNATCAGCTNAGGTTAG.AA 459
327 nHlStrAlaGluGlyThrGlyThrAsnLysLysValAlaLysArgAsnA 344
458 CCCCATTGCANAGAGAGCGGCACACAGAAG...TGCCACGCAAT 412
344 lAlaLysAsnMetLeuGluIleLeuGlyPheLysValProGlnArgGln 360
411 GCAGCGAGANCATGTGAAGATCCTGGGTTCAAGTCCCGCGCGCGAG 362
361 ProThrLysProAlaLeuLysSerGluGluLysThrProIleLysLysP 377
361 CCACC.AAACCGGCACTCAAGTCAGAGAGACACCCCTAAGAAGACC 313
377 oGlyAspGlyArgLysValThrPheheGluProGlySerGlyAspGluA 394
312 AGGGATGGAGAGAAAGTAACCTTTTGAACCTGCTCGGGGATGAAA 263
394 sngLyThrSeranlySgluAspGluPheArgMetProTyrIleuSerHis 410
262 ATGGACTAGTAATAAGAGATGAGTTCAGGATGCTTATCTAAGTCAT 213
411 GlnGlnLeuProAlaGlyIleLeuProMetValProGluValAlaGlnAl 427
212 CAGCAGCTGCTCTGGAATTTCTCCATGTGTGCCGAGGTGCCACAGGC 163
427 aValGlyValSerGlnGlyHisHisThrLysAspPheThrArgAlaAla 444
162 TGTAGAGATTAGTCAAGACATCACACAAAGATTTTACAGGCGCAGCTC 113
444 rObAsnProAlaLysAlaThrValThrAlaMetIleAlaArgGluLeuLeu 460
112 CGAATCCTGCCAAGGCCAGCGTAACGTCATGTATACCCGAGGTTGTG 63
461 TyrGlyGlyThrSerProThrAlaGluThrIleLeuLysAsnLysLeu 477
62 TATGGGGGACCTCGCCACAGCCGAGACCATTTTAAGATATACATCTC 13
477 rSerGlyHis 480
12 TTCAGGCCAC 3
seq_name: /SID6/gcdata/geneseq/geneseqn/NA2000.DAT:280269
seq_documentation_block:
ID 280269 standard; cDNA; 727 BP.

```

```

XX 280269;
AC
XX
XX 07-APR-2000 (first entry)
DT
XX
XX Human colon cancer cell line SW480 cDNA clone SEQ ID NO:353.
DE
XX
XX Human; gene expression product; diagnosis; tumour; colon cancer;
KM colorectal adenocarcinoma; cell line SW480; cell proliferation;
KM cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
KM hyperplasia; ds.
XX
XX Homo sapiens.
OS
XX
XX WO964576-A2.
PN
XX
XX 16-DEC-1999.
PD
XX
XX 09-JUN-1999; 99WO-1B01062.
PF
XX
XX 10-JUN-1998; 98US-0088801.
PR
XX
XX (FARB ) BAYER CORP.
PA
XX
XX Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catino TJ, Derli A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;
XX
XX WPI: 2000-087220/07.
DR
XX
XX Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer
XX
XX Claim 15; Page 273; 469pp; English.
PS
XX
XX 279917 to 280766 represent double stranded cDNA clones isolated from the
CC human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g. carcinomas
CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used
CC for developing agents for the diagnosis and treatment of disorders
CC involving unwanted cell proliferation, such as neoplasia, dysplasia or
CC hyperplasia.
XX
XX Sequence 727 BP; 194 A; 153 C; 191 G; 149 T; 40 other;
SQ

```

```

alignment_scores:
  Quality: 691.50      Length: 155
  Ratio: 4.704        Gaps: 2
Percent Similarity: 94.839 Percent Identity: 90.323

```

```

alignment_block:
US-09-316-048-6 x 280269 ..

```

```

Align seg 1/1 to: 280269 from: 1 to: 727

```

```

116 TyrPheYrPProPheProValProProLeuLeuYrGlnValGluLeuSe 132
|||||
3 TACTTTTACCATTTCAGTTCACCTTTACTTTATCAAGTGAACCTTTC 52
132 TValGlyGlnGlnPheAsnGlyLysGlyLysThrArgGlnAlaAlaL 149
|||||
53 TGTGGAGAGACAGCAATTATGCAAGAAAGACAAGACAGCGCTGCGA 102
149 ySHISAspAlaAlaAlaAlaAlaLeuArgGlnLeuGlnAsnGluProLeu 165
|||||
103 AACACGAGTGTGTCGCAAGCGTTGAGATCTGCAGAAATGAGCCCTG 152
166 ProGluArgLeuGluValAsnGlyArgGluSerGluGluAsnLeuAs 182

```

```

|||||
153 CCAGAGAGCGTGTGATGTAAGAGAAATCCGAAAGAAATCTCAA 202
182 nLYSeRGlutIleSerGlnValPheGluIleAlaLeuLysArgAsnLeuP 199
|||||
203 TAAATCTGAATTAAGTCAAGCTTTGAGATTGCACCTTAACGGAACCTTGC 252
199 roValAsnPheGluValAlaAlaArgGluSerGlyProProHisMetLysAsn 215
253 CTGTGAATTTCGAGTGCGCCGGAGAGATGGCCACCCACATGAGAAGAC 302
216 PheValThrLysValIleSerValGlyGluPheValGlyGluGlyGly 232
303 TTTGTGACCAAGCTTTCGTTGGGAGATTGTGGGGAAGGTAAGGGA 352
232 sSerLysLysIleSerLysLysAsnAlaAlaIleAlaValLeuGlu 248
353 AAGCAGAGAGATTTCAAGAAAATGCCCGCATAGCTGNTCTTGAGAGAG 402
249 LeuLysLysLeuProProLeuProAlaVal...GluArgValLysProAr 264
403 CTGAAGAAAGTACCGCCCTGCTTGNATTGACCGAAGTAAAGGCTTAN 452
264 gTLeLysLysLys 268
453 AATCCAAATGAAA 465
seq_name: /SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:217411
seq_documentation_block:
ID 217411 standard; cDNA; 769 BP.
AC 217411;
XX
DT 12-OCT-1999 (first entry)
XX
DE Human gene expression product cDNA sequence SEQ ID NO:4884.
XX
XX Human; gene; gene expression product; diagnosis; therapy; probe;
KM detection; mapping; tissue typing; profiling; forensic; cancer;
KM genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
XX Homo sapiens.
OS
XX
XX WO9938972-A2.
PN
XX
XX 05-AUG-1999.
PD
XX
XX 28-JAN-1999; 99WO-US01619.
PF
XX
XX 03-APR-1998; 98US-0080666.
PR 28-JAN-1998; 98US-0072810.
PR 24-FEB-1998; 98US-0075954.
PR 31-MAR-1998; 98US-0080114.
PR 03-APR-1998; 98US-0080515.
XX
XX (CHIR ) CHIRON CORP.
PA
XX (HYSE-) HYSEQ INC.
XX
XX CrkVenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innt MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stachne-Crain B, Sudduth-Klinger J, Williams LT;
XX
XX WPI: 1999-494092/41.
XX
XX Novel human genes and their expression products which are
PT differentially expressed in different cell types
XX
XX Claim 1; Page 2322; 2479pp; English.
XX
XX The present invention describes a library of human polynucleotides
CC

```

comprising the sequences given in Z12532 to Z17779. Also described is a method of detecting differentially expressed genes correlated with the cancerous state of a mammalian cell, comprising detecting at least one differentially expressed gene product in a test sample from a cell suspected of being cancerous, where the gene product is encoded by one of the 5248 polynucleotide sequences given in Z12532 to Z17779. The polynucleotides can be used as a source of primers and probes, which can be used for a variety of purpose, e.g. detection of expression levels, mapping, tissue typing or profiling, forensics, genetic analysis and detection of polymorphisms. Polypeptides encoded by the polynucleotides can be used for raising antibodies for experimental, diagnostic and therapeutic purposes. The polynucleotides may also be used to construct arrays for diagnostics (which may be used to determine function of an encoded protein); and to detect differences in expression levels between two cells (e.g. to identify abnormal or diseased tissue in a human, to identify a genetic predisposition or susceptibility to a disease such as cancer). The polynucleotides of the invention are especially used in the diagnosis, prognosis and management of colorectal cancer, breast cancer, and lung cancer. The polynucleotides can also be used to screen for peptide analogues and antagonists.

Sequence 769 BP; 176 A; 214 C; 135 G; 188 T; 56 other:

alignment\_scores:  
Quality: 418.00 Length: 126  
Ratio: 3.835 Gaps: 5  
Percent Similarity: 86.508 Percent Identity: 78.571

alignment\_block:

US-09-316-048-6 x Z17411 ..

Align seg 1/1 to: Z17411 from: 1 to: 769

```

1 MetSerGlnValGlnValGlnValGlnAsnProSerAlaAlaLeuSerG1 17
|||||
277 ATGCTCAGATTCAAGTTCAGAACCCATCTGCTGCTCTCTCAGG 326
17 yserGlnIleuAsnLysAsnGlnSerLeuSerGlnProLeuMetS 34
|||||
327 GACCCAAATACGACACAGACAGTCTTCTCTCACACCCCTTGATGA 376
34 erlleProSerThrThrSerSerLeuProSerGlnAsnAlaGlyArgPro 50
377 GATTCCTCTTACTACTAGCTCTGCTGCCCTGAAATAGAGTAGAGCC 426
51 lleGlnAsnSerAlaLeuProSerAlaSerlleThrSerThrSerAla 67
|||||
427 ATTCAAACACTGCTGTTTACCTCTGCATCTATACATCCNACCAGTG 476
67 laAlaGlnSerlleThrProThrValGlnLeuAsnAlaLeuGlyMetLys 83
|||||
477 NTGCAAAAGCATTAACCTTACTGTAGAACTTAATGCTGGGATGAAAC 526
84 LeuGlyLysLysProMetLysLysProValAspProTyrSerArgMet 99
|||||
527 TTGGAAAAAACCAATGTTAAGCCNTGTTGAACCTTACTCTCGGGATG 576
100 GlnSerThrTyrAsnTyrAsn.....MetArgGlyGlyAlaTyrPr 113
|||||
577 CAGNCACCTTAATACTACCAACATGAGAGNANGAAGAGGTTAATCC 626
113 oProArgTyr...PheTyrPro 119
|||||
627 CCCNAGGNNACTTTTNNCCA 648

```

seq\_name: /SID6/gcgdata/geneseq/NA1999.DAT: X24067

seq\_documentation\_block:

ID X24067 standard: DNA: 1904 BP.

AC X24067;

XX

DT 30-JUN-1999 (first entry)

XX Human regulatory molecule HRM-9 DNA.

KW Human regulatory molecule; HRM-9; cytostatic activity; immune modulator;  
KW transcription factor; enhancer; cell proliferation stimulation; cancer;  
KW treatment; microarray; detection; diagnosis; cell proliferation disease;  
KW leukemia; lymphoma; myeloma; adenocarcinoma; sarcoma; bladder; bone;  
KW brain; lung; liver; ovary; skin; teratocarcinoma; immune response;  
KW allergy; asthma; diabetes; multiple sclerosis; Grave's disease;  
KW myasthenia gravis; ss.

XX Homo sapiens.

XX WO9915658-A2.

XX 01-APR-1999.

XX 22-SEP-1998; 98WO-US19839.

XX 23-SEP-1997; 97US-0933750.

XX (INCY-) INCYTE PHARM INC.

XX Au-Young J, Bandman O, Guegler KJ, Hillman JL, Lal P;

XX Shah P, Yue H;

XX WPI.1999-254710/21.

XX P-PSDB: W93953.

XX New human regulatory molecules

XX Claim 5; Page 75; 76pp; English.

CC This invention describes novel human regulatory molecules (HRM) which  
CC have cytostatic activity and act as immune modulators, transcription  
CC factors or enhancers. The HRMs can be used to stimulate cell  
CC proliferation. Antagonists and agonists of the proteins of the invention  
CC can be used to treat cancer. The encoding nucleic acids can be used in  
CC microarrays to detect polynucleotides (and their expression levels) that  
CC encode HRMs in a biological sample. The HRMs and microarrays can be used  
CC to diagnose, treat or prevent cell proliferation diseases especially cancer,  
CC e.g. leukemia, lymphoma, myeloma, adenocarcinoma, sarcoma, cancer of e.g.  
CC bladder, bone, brain, lung, liver, ovary, skin, etc, teratocarcinoma, or  
CC to treat or prevent immune responses e.g. allergies, asthma, diabetes,  
CC multiple sclerosis, Grave's disease or myasthenia gravis.

XX Sequence 1904 BP; 576 A; 401 C; 385 G; 540 T; 2 other:

alignment\_scores:  
Quality: 218.00 Length: 490  
Ratio: 1.033 Gaps: 15  
Percent Similarity: 43.061 Percent Identity: 22.041

alignment\_block:

US-09-316-048-6 x X24067 ..

Align seg 1/1 to: X24067 from: 1 to: 1904

```

134 GlyGlyGlnGlnIlePheAsnGlyLysGlyLysThrArgGlnAlaAlaLysH1 150
|||||
17 GGTGGGGGGCGCTGANNCTCTCCGGGGGACACGACCGGGAGCATGCA 66
150 saspaAlaAlaAlaLysAlaLeuArgIleLeuGlnAsnGluProLeuPro. 166
1 .....
67 CGTGTGTCACACGACTTCACTGAGCTTGGAACAGAACTAATCTCCCTC 116
1 .....
166 .....
117 CATCTCCAACTTGCTTGACAAAGCTGCTCTTCTCCAGCGGTGAGCCG 166
167 .....GluArgLeuGluValAsnGlyAr 174

```

```

167 TCCTTCTGCCATGTCCAGACGACGCCGCGCGCCGCGCT 216
174 ggluserglugluasnlenu..... 181
217 GGAGCCGAGACAGTGGACCTTCAGTTGGGAGATGATACAGCTA 266
182 ....AsnLysSerGluIleSerGluValPheGluIleAlaLeuLys... 195
267 AGCAGGAGAAACACCGATTCAGGTATTACAGAAATACGCATGAAGACC 316
196 ArgAsnLeuProValAsnPheGluValAlaArgLysSerGlyProProH 212
317 AAGAACATCCACGT...TATGAATGGAAGATCTCATGTGCAGAAATACA 363
212 smetLysAsnPheValThrLysValSerValGlyGluPheValGlyGlu 229
364 CGTGCCCACTTCACCTTCAGAGTACCGTTGGTGAATACCTGCACAG 413
229 LysGluLysSerLysLysIleSerLysLysAsnAlaAlaAlaAlaVal 245
414 GTGAGGTACAACTAGAACTGGCGAAACATAGAGCTGCAGAGCTGCC 463
246 LeuGluGluLeuLysLysLeuProProLeuProAlaValGluArgVal 262
464 ATAAACATTTTGA... 483
262 sProAlaGlyLeuLysLysThrLysProIleValLysProGlnThrSer 279
484 TGCAGATATTGCTTTCAGCTCCAGCCCTTAATGCTGCACCTTCCA 533
279 roGluArgGlyGlnGlyIleAsnProIleSerArgLeuAlaGlnIleGln 295
534 AGCAACCAAGAACACGCTTAATCCTATGTTGTTACAGGATTCGCT 583
296 GlnAlaLysLysGluLysGluProGluThrLeuLeuThrGluArgG1 312
584 ATTCATCATGCTGAGACTTCTGATATACCTTCCAGAGGAGGAGG 633
312 yLeuProArgArgArgGluPheValMetGlnValLysValGlyAsnHisT 329
634 ACCTGTCATAAGAGAGATATCTACAAATTTGAGGCTGAGCTTGA 683
329 hrAlaGluGlyThrGlyThrAsnLysLysValAlaLysArgAsnAlaAla 345
684 TGGAACTGGAAGGGGCGATCAAAAAAGCAAGCAAAAGGATGCTGCT 733
346 GluAsnMetLeuGluIleLeuGlyPheLysValProGlnArgGlnProH 362
734 GAGAAATTTCTT..... 745
362 rLysProAlaLeuLysSerGluLysThrProIleLysLysProGlyA 379
745 ..... 745
379 spGlyArgLysValThrPhePheGluProGlySerGlyAspGluAsnGly 395
746 ....GCCAAATTAGTAATATTCTCA.....GAGAACAC 778
396 ThrSerAsnLysGluAspGluPheArgMetProTyrLeuSerHisGlnG1 412
779 ATTTCT..... 784
412 nLeuProAlaGlyIleLeuProMetValProGluValAlaGlnAlaValG 429
785 .....TTAACAAATGTAAGTAG 800
429 LysValSerGlnGly.....HisHisThrLysAspPheThrArgAla 442
801 GACATTTTGAAGTACTGATTCCTTGAAGAT..... 838
443 AlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArgGluLe 459

```

```

838 ..... 838
459 uLeuTyrGlyGlyThrSerProThrAlaGlu...ThrIleLeuLysAsnA 475
839 .....TTCCTGGTGAAGATCACTTACTGAAAGAA 872
475 snIleSerSerGlyHisValProHisGlyProLeuThrArgProSerGlu 491
873 GCCCTCTTACT...ATTCCAAT.....ACAGATTACATCAG 907
492 GlnLeuAspTyrLeuSerArgValGlnGlyPheGlnValGluTyrLysAs 508
908 CTGCTTGTAAATGGCAAGAACAGGTTTAAATTAATATATTTGGA 957
508 pPheProLysAsnAsnLysAsn...GluPheValSerLeuIleAsnCysS 524
958 TATATGATGAACGTAGCGCCCAATGACAAATATCAATGCTTCGTAACGT 1007
524 eSerGlnProProLeuIleSerHisGlyIleGlyLysAspValGluSer 540
1008 CCACACAGCCCATCAGATCTGTCATGCTGCGGTATCTCCTGGCAAT 1057
541 CysHisAspMetAlaAlaLeuAsnIleLeuLysLeu..... 553
1058 GCACAAAGTATGCAGCTCACAAATGCTTGCAGATTAAAGATAATAGC 1107
554 .....SerGluLeuAspGlnGlnSerThrGluMetP 564
1108 AGAAAGAAAGTAATCTGAGCACTTAATAAATCTTACGTACACATA 1157
564 roArgThrGlyAsnGlyPro 570
1158 AAAAGTCCCTCCTGCGCCC 1177

seq_name: /STD6/gcgdata/geneseq/geneseqn/NA2000.DAT.Z52512
seq_documentation_block:
ID 252512 standard; cDNA; 1396 BP.
XX
AC 252512;
XX
DT 29-FEB-2000 (first entry)
XX
DE Human secreted protein clone yj10_1 nucleotide sequence SEQ ID NO:75.
XX
KW Human; secreted protein; immunostimulatory; haemostatic; cytokine;
KW proliferative; differentiative; chemotactic; chemokine; vaccine;
KW thrombolytic; antiinflammatory; cytostatic; immunosuppressive;
KW gene therapy; ss.
XX
OS Homo sapiens.
XX
PN W0958642-A2.
XX
PD 18-NOV-1999.
XX
PE 14-MAY-1999; 99MO-US10843.
XX
PR 14-MAY-1998; 98US-0085472.
PR 17-AUG-1998; 98US-0096824.
PR 11-SEP-1998; 98US-0099843.
PR 11-SEP-1998; 98US-0099950.
PR 15-SEP-1998; 98US-0100424.
PR 29-SEP-1998; 98US-0102329.
PR 09-OCT-1998; 98US-0103615.
PR 11-DEC-1998; 98US-0111799.
PR 14-DEC-1998; 98US-0112159.
PR 31-DEC-1998; 98US-0114415.
PR 10-FEB-1999; 98US-0248059.
PR 06-APR-1999; 98US-0287150.
PR 13-MAY-1999; 99US-0311021.
XX
(GEMV ) GENETICS INST INC.

```

```

XX  Wong GG, Clark HF, Fechtel K, Agostino MJ;
PI  WPI: 2000-053095/04.
XX  P-PSDB: Y73427.
DR
XX  Novel polynucleotides and proteins having biological activities which
PI  make them suitable for treating, preventing or ameliorating medical
XX  conditions in humans or animals -
PS  Claim 84; Page 631-632; 730pp; English.
XX
XX  The present invention describes human secreted proteins encoded by
CC  polynucleotides obtained from adult testes, foetal brain, adult brain,
CC  brain (foetal and adult), foetal kidney, adult spleen, and adult thymus
CC  CDNA libraries. The polynucleotides and proteins are predicted to have
CC  biological activities which would make them suitable for treating,
CC  preventing or ameliorating medical conditions in humans and animals.
CC  Suggested activities include nutritional activity, cytokine and cell
CC  proliferation/differentiation activity, immune stimulating (e.g. as
CC  vaccines) or suppressing activity, haematopoiesis regulating activity,
CC  tissue growth activity, activin/inhibin activity, chemotactic/
CC  chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC  ligand activity, anti-inflammatory activity, cadherin/tumour invasion
CC  suppressor activity, and tumour inhibition activity. The polynucleotides
CC  are also stated to be useful for gene therapy. Therapeutic compositions
CC  are also presently valuable for veterinary applications. 252475 to 252581
CC  encode human secreted proteins, and Y73390 to Y73500 represent human
CC  secreted proteins, given in the present invention.
XX
XX  Sequence 1396 BP; 439 A; 313 C; 292 G; 352 T; 0 other:
SQ
XX
XX  alignment_scores:
XX      Quality: 211.50      Length: 404
XX      Ratio: 1.143      Gaps: 13
XX      Percent Similarity: 45.792      Percent Identity: 23.267
XX
XX  alignment_block:
XX  US-09-316-048-6 x 252512 ..
XX
XX  Align seg 1/1 to: 252512 from: 1 to: 1396
XX
XX  183 LysSerGluIleSerGlnValPheGluIleAlaLeuLys...ArgAsnLe 198
XX  |||:: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
XX  205 AAACACCCGATTTCAGTATTACACGAATACGGCATGACAGACCAAGAAAT 254
XX
XX  198 uProValAsnPheGluValAlaArgGluSerGlyProProHisMetLysA 215
XX  ::|||::: |||::: |||::: |||::: |||::: |||::: |||:::
XX  225 CCCAGTT...TATGAATGTGAAGATCTGATGTGCAAAATACACGTGCCA 301
XX
XX  215 snPheValThrLysValSerValGlyLupPheValGlyGlyGlyGly 231
XX  ::|||::: |||::: |||::: |||::: |||::: |||::: |||:::
XX  302 CTTTCACCTTCAGATACCGCTGTGTCACATACCTGCACAGGTGAAGGT 351
XX
XX  232 LysSerLysLysIleSerLysLysAsnAlaAlaIleAlaValLeuGlu 248
XX  |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
XX  352 ACAAGTAAGAGCTGGCGAAACATAGAGCTGCAGAGGCTGCCATTAACAT 401
XX
XX  248 uLeuLysLysLeuProProLeuProAlaValGluArgValLysProArg 265
XX  |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
XX  402 TTGTGAAA.....GCCAATGCACAGTA 421
XX
XX  265 IeLysLysLysThrLysProIleValLysProGlnThrSerProGluTyr 281
XX  ::|||::: |||::: |||::: |||::: |||::: |||::: |||:::
XX  422 TTGTGTTTCAGATTCTGACCCCTTAATGCGTGCACCTTCACAGCAACA 471
XX
XX  282 GlyGlyGlyIleAsnProIleSerArgLeuAlaGlnIleGlnGlnAla 298
XX  ::|||::: |||::: |||::: |||::: |||::: |||::: |||:::
XX  472 AAGACACAGCTTAATCCTATTGGTTCATTACAGGAATGGCATTCATCA 521
XX  sLysGluLysGluProGluTyrThrLeuLeuThrGluArgGlyLeuProA 315
XX  ::|||:: |||::|||::|||::|||::|||::|||::|||::|||::

```

```

522 TGGCTGAGACATTCTGAATATACCTTTCCAGAGGAGGACCTGCTC 571
315 rGArArgGluPheValMetGlnValLysValGlyLysnHisThrAlaGlu 331
::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
572 ATAGAGAGAGATATACATATTGCAAGCTAGAGCTATTTATGAAACT 621
332 GLYThrGlyThrAsnLysLysValAlaLysArgAsnAlaIleLysnMe 348
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
622 GGAAGGGGGGCATCAAAAAAGCAAGCAAAAGAGATGCTCTCAGAAAT 671
348 tLeuGluIleLeuGlyPheLysValProGlnArgGlnProThrLysProA 365
::|||::
672 TCTT..... 675
365 IeLeuLysSerGluGluLysThrProIleLysLysProGluAspGlyArg 381
676 .....GCC 678
382 LysValThrPhePheGluProGlySerGlyAspGluAsnGlyThrSerAs 398
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
679 AATTTTAGTAAATATTCTCTCA.....GAGACCAATTTCT.. 714
398 nLysGluAspGluPheArgMetProTyrLeuSerHisGlnLeuProA 415
714 ..... 714
415 IeLysIleLeuProMetValProGluValAlaGlnAlaValGlyLysSer 431
::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
715 .....TTAACAATGATGAGAGACATTTCT 738
432 GlnGly.....HisHisThrLysAspPheThrArgAlaIleProAs 445
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
739 TTAGCATGTACTTGGCATTCCTTGAGAGAAAT..... 768
445 nProAlaLysAlaThrValThrAlaMetIleAlaArgGluLeuLeuTyrG 462
768 ..... 768
462 LysGlyThrSerProThrAlaGlu...ThrIleLeuLysAsnAsnIleSer 477
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
769 .....TCTCTGGTGAAGATCAACTTACTGAAAGAAAGACCTCTCT 810
478 SerGlyHisValProHisGlyProLeuThrArgProSerGluGlnLeuAs 494
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
811 AGT.....ATTCCAAT.....ACAGATTACATCAGCTGCTTAG 845
494 pTyrLeuSerArgValGlnGlyPheGlnValGluTyrLysAspPheProL 511
::|||::: |||::: |||::: |||::: |||::: |||::: |||:::
846 TGAATTTGCCAAGCAAGAGGTTTATATATACATATTGGAATATAGATG 895
511 yAsnAsnLysAsn...GluPheValSerLeuIleAsnLysSerSerGln 526
::|||::: |||::: |||::: |||::: |||::: |||::: |||:::
896 AACTGAGCGGCAATGACATATATCAATGTCTTCTGAACGTCCACACG 945
527 ProProLeuIleSerHisGlyIleGlyLysAspValGlyIleSerCysHisAs 543
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
946 CCCATACAGCTGTGTATGCTCCGATATCTCTGTGGCAATGACAAAG 995
543 pMetAlaAlaLeuAsnIleLeuLysLeuLeu..... 553
::|||::: |||::: |||::: |||::: |||::: |||::: |||:::
996 TGATGCAGCTCACCAATGCTTTCAGATATTAAAGATATAGCAGAAAGAA 1045
554 .....SerGluLeuAspGlnGlnSerThrGluMetProArgThr 566
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
1046 AGTAAATCTGGACCAACTTAAAAAATCTTCACTAGCACACATAAAGATTC 1095
567 GlyAsnGlyPro 570
::|||::: |||::: |||::: |||::: |||::: |||::: |||:::
1096 CCTCTGGCCCC 1107

```









51 AsnGlyLysGlyMetArgProProValLysHisAspAlaProAlaAr 67  
 476 AATGGAAAGAAAGATGAGACACCCGTGAACACATGACCATGCGCCG 525  
 67 gAlaLeuArgThrLeuGlnSerGluProLeuProGluArgLeuGluValA 84  
 526 TGCCTGAGAGACTGTCAGAGTGAACCCCTGCCAGAAAGTTGAGAGTAA 575  
 84 snGlyArgGluAlaGluGluLysAsnLysAsnLysSerGluLysSerGln 100  
 576 ATGGAAGAAGAACGAGAGAAACCTCAATTAATCGAAATTAAGCCAA 625  
 101 ValPheGluLeuLeuLeuLysArgAsnLeuProValAsnPheGluValA 117  
 626 GTGTTGAATTTGCGCTGAAGCGGAATTTGCTGTAATTTTGAGGTGGC 675  
 117 aArgGluSerGlyProProHisMetLysAsnPheValThrArgValSerV 134  
 676 CCGGAGAGTGGCCACACACATGAAGAACTTTGTGACCGAGGTTCAG 725  
 134 aLgylGluPheValGlyGluGlyGlyLysSerLysLysLysLys 150  
 726 TTGGGGATTTGTAGGGAGAGAGAGAGAAAGCAAGAAAGATCTCCAG 775  
 151 LysAsnAlaAlaArgAlaValLeuGluGlnLeuArgArgLeuProPhe 167  
 776 AGAAATCGCGGAGGCGCTTTCTGACAGCTTGAAGCGTCCACCCCT 825  
 167 uProAlaValAlaArgValLysProArgLysLysLysSerGlnProT 184  
 826 CCCGCTGTGGAGAGCTGAAGCCCAATCAAGAAAGAAAGTACGCCCA 875  
 184 hrcLysLysThrAlaProAspTyrGlyGlnGlyMetAsnProLysSerArg 200  
 876 CTTGCAAGACACCCCGGATTATGCGCAAGGATGAATCTATTATGAGA 925  
 201 LeuAlaGlnLeuGlnAlaLysLysGluLysGluProGluTyrMetLe 217  
 926 CTTGCAAGATCCAGCAGCAAAAAAGAGAGAGAGAGAGATCATGCT 975  
 217 uLeuThrGluArgGlyLeuProArgArgGluPheValMetGlnValL 234  
 976 CTTACGAGAACGAGTCTCCACCTCCAGAGGAGTTGTGTGACAGGTAA 1025  
 234 yValGlyHisThrAlaGluGluValGlyThrAsnLysValAla 250  
 1026 AGGTGGGCACTCACACTGCAGAGAGATGGTACCAATTAAGAGGTGCC 1075  
 251 LysArgAsnAlaAlaGluAsnMetLeuGluLeuGlyPheLysValPr 267  
 1076 AAGCGTAATGCTGCTGAGAACATCTGAGATCCTGGGTTCAAAAGTTCC 1125  
 267 ogAlaGlnProAlaLysProAlaLeuLysSerGluGluLysThrProV 284  
 1126 CCAAGGCCACCTGCCAAGCAGCACAATCAAGAGAGAGAGAGAGTCCAG 1175  
 284 aLysLysProGlyAspGlyArgLysValThrPhePheGluProSerPro 300  
 1176 TAAAGAAACCAAGAGAGAGAGAGAAAGTAACTTTTGAACCTAGACCT 1225  
 301 GlyAspGluAsnGlyThrSerAsnLysAspGluGluPheArgMetProTy 317  
 1226 GGGGATGAATGAGAACTAGTAACAGAGAGAGAGAGTTCAGAGATGCCCTTA 1275  
 317 rLeuSerHisGlnGlnLeuProAlaGlyLysLeuProMetValProGluV 334  
 1276 TCTTAGCCATCAGCAGCTCCAGCTGGAATTTCTCCATGGTGGCGGAG 1325  
 334 aAlaGlnAlaValAlaGlyValSerGlnGlyHisThrLysAspPheThr 350  
 1326 TTGCCCGAGCGCTGTGGGTTAGTCAAGAGACCAACCAAGATTTTACC 1375  
 351 ArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetLeuAlaAr 367

1376 AGGCGACCTCCAAATCCTGCCAAGCAAGCTACTGCATGATAGCCCG 1425  
 367 gGluLeuLeuTyrGlyLysThrSerProThrAlaGluThrLeuLys 384  
 1426 AGAGTTGTGTGACGGGGACCTGCGCCACAGCCGAGACCAATTTTAAAGA 1475  
 384 eAsnLysSerSerGlyHisValProHisGlyProArgThrArgProSer 400  
 1476 GTAACATCTCTTCAGGCGACGTACCCATGAGACCTGCGACAGACCTCT 1525  
 401 GluGlnLeuTyrTyrLeuSerArgAlaGlnGlyPheGlnValGlyTyrL 417  
 1526 GAGCAACTGTACTACTCTTCAGAGCCAGGATTCAGAGTTGAATCAAA 1575  
 417 sAspPheProLysAsnAsnLysAsnGluCysValSerLeuLysnYss 434  
 1576 AGATTTTCCCAAGAACAAAGAACGAGTGTATCTCTCATCAACTGCT 1625  
 434 eSerGlnProProLeuValSerHisGlyLysLysAspValGluSer 450  
 1626 CTTACAGCCGCTCTCTGTCATGATGATGCGCAAGAGATGTGAGTCC 1675  
 451 CysHisAspMetAlaAlaLeuAsnLeuLeuLysLeuLeuSerGluLeuAs 467  
 1676 TGTCATGATATGCTGCTGCTGACATGATTTAAAGCTGCTGTGATTGGA 1725  
 467 pGlnGlnSerThrGluMetProArgThrGlyAsnGlyProValSerAlaC 484  
 1726 CCAACAGAGCAGACAGATGCCAAGAGAGAAATGAGACCAAGTTTCACGT 1775  
 484 ysgLysArgCys 487  
 1776 GCGGAGAGTGC 1786  
 seq\_name: gb\_ro:AF290989  
 seq\_documentation\_block: 3041 bp mRNA ROD 09-AUG-2000  
 LOCUS AF290989 Rattus norvegicus stauufen isoform stau-16 (stau) mRNA, complete  
 DEFINITION cds, alternatively spliced.  
 ACCESSION AF290989  
 VERSION AF290989.1 GI:9755162  
 KEYWORDS  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 REFERENCE 1 (bases 1 to 3041)  
 Monshausen, M., Putz, U., Rehbein, M., Schweizer, M.,  
 Desroselliers, L., Kuhl, D., Richter, D. and Kindler, S.  
 Two rat brain stauufen isoforms differentially bind RNA  
 JOURNAL Unpublished  
 TITLE 2 (bases 1 to 3041)  
 REFERENCE Monshausen, M. and Kindler, S.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-JUN-2000) Institute for Cell Biochemistry and  
 Clinical Neurobiology, University of Hamburg, University Hospital  
 Eppendorf, Sudefeldstr. 24, Hamburg D-22529, Germany  
 FEATURES  
 source  
 1..3041  
 /organism="Rattus norvegicus"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /tissue\_type="brain"  
 /dev\_stage="adult"  
 1..3041  
 /gene="Stau"  
 227..1696  
 /gene="Stau"  
 /note="RNA-binding protein; similar to Drosophila stauufen;  
 contains four RNA-binding domains; rStau-16"

```
/codon_start=1
/product="staufen isoform Stau-16"
/product_id="AAF98177.1"
/db_xref="GI:9755163"
/translation="MKKPYDPHSRMOSTSYSGKRGAPPPRYFPYPPPLLYQVLS
VGGQPNKGKMRPPYKHDTARALTLTLOSEPLPEKLEVNGRSEENLNKSETISOV
ELALKNLPLVNEVARESGPPHKNFVTRVSGVEGEEBSKKISKKNAARAVLEO
LRLPLPAVEREVKPRIKKSOPTCKLQTPADYGQGNPISRLAQIOAKKEKEPEYM
LTERLPRRREVMQVKGNHTAEGAGINKKVKARNAENMLEIGFYVPOPOPAK
AKSEKTPVKKPDGGRKRYTFEPSPDNGSNKDEPRMPLYSHOOLPAQILPMVP
EYAAQVSGQSHHTKDEFAARAAPAKATYTAIARLLIGGTSPTRETIKSISSGH
VPHGPRTPSEQLYIUSRAQFOVERKDFPKNNKNCSVLINCSQPLVSHGIGDV
ESCHDMAALNIKLSELDQSTEMPRGTGNPVSACRC"
polyA_signal
/gene="Stau"
```

BASE COUNT 814 a 771 c 747 g 707 t 2 others  
ORIGIN

## alignment\_scores:

Quality: 2512.00 Length: 489  
Ratio: 5.169 Gaps: 1  
Percent Similarity: 99.387 Percent Identity: 98.364

## alignment\_block:

US-09-316-048-11 x AF290989 ..

Align seg 1/1 to: AF290989 from: 1 to: 3041

```
1 MetTyrLysProValAspProHisSerArgMetGlnSerThrTyrSerTyr 17
|||||
227 ATGTATTAAGCCCGTGGACCCCTCTCGATGTCAGTCCACCTACAGCTA 276
17 rGLeuMetArgGlyAlaTyrProProArgTyrPheTyrProPheProy 34
|||||
277 TGGCATGCGTGAGGTGCTATCCGCCAGATACTTTAACCATTTTCAG 326
34 aLProProLeuLeuTyrGlnValAlaGluLeuSerValGlyGlnGlnHe 50
|||||
327 TTCCACCGTTACTCTACCAAGTTGACGCTCTCCGTGGAGGAGCAGAGTTT 376
51 AsnGlyLysGlyLysMetArgProProValLysHisAspAlaProAlaAr 67
|||||
377 AATGGGAAGGAAAGATGAGACCACTGTGAACATGATGCCACATGCCCG 426
67 gAlaLeuArgThrLeuGlnSerGluProLeuProGluArgLeuGluValA 84
|||||
427 AGCCCTGAGGACTCTGCAGAGTGAAGCCCTACACAGAGGCTAGAGGTAA 476
84 snGlyArgGluAlaGluGlnGluAsnLeuAsnLysSerGluLleSerGln 100
|||||
477 ATGGAAAGAAATCAGAAAGAAAACCTTAATAATTCAGAAATTAAGCCAA 526
101 ValPheGluIleAlaLeuLysArgAsnLeuProValAsnPheGluValAl 117
|||||
527 GTGTTTGAGATGGCGTGAAGCGGAATTTGCCCTGTGAATTTTGAAGTGC 576
117 aArgGluSerGlyProProHisMetLysAsnPheValThrArgValSerV 134
|||||
577 CCGGAGAGGTGGCCGCCACATGAAGAACTTTGTGACACAGGGGTTTCAG 626
134 aLgLyGluPheValGlyGlnGlyGluGlyLysSerLysLysLleSerLys 150
|||||
627 TTGGGGAGTTTGTAGGGGAAGGAAGGAACCAAGAAAGATCTCCCAAG 676
151 LysAsnAlaAlaArgAlaValLeuGlnLeuArgArgLeuProProLe 167
677 AAGATGGCGGAAGGGCGCTGTGAGAGAGCTAAGAGAGCTGCCACCTCT 726
167 uProAlaValGluArgValLysProArgLleLysLysLysSerGlnProT 184
727 ACCTGCTGTGGAGGAGTGAAGCCAGAAATCAGAAAGAAAGTACGCCCA 776
```

```
184 hrCysLys.....ThrAlaProAspTyrGlyGlnGlyMetAsnProIle 198
|||||
777 CCTGCAAGCTACAGACTGCCCCCGACTATGGCCAGAGGATGTAATCTTAT 826
199 SerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTyr 215
|||||
827 AGCAACACTGGCACAGATCCAGCAGCAAAAAAGGGAAGGAGCAGAGTA 876
215 rMetLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMetG 232
|||||
877 CATGCTCTTACAGACAGAGCTTCTCCACGTGCGAGGGAGTTTGATGATC 926
232 InValLysValGlyHisThrAlaGluGlyValGlyThrAsnLysLys 248
|||||
927 AGGTCAAGGTTGGGATCATCACACCGCAGAAAGAGAGTAAACCAATAAG 976
249 ValAlaLysArgAsnAlaAlaGluAsnMetLeuGluLleLeuGlyPheL 265
|||||
977 GTGGCCAAAGCCGAATGCTGTGAGAACATGCTGAGATCTGGGGTTCAA 1026
265 sValProGlnAlaGlnProAlaLysProAlaLeuLysSerGluGluLysT 282
|||||
1027 AGTCCCAAGGCGCAGCCTGCGCAGCCAGCTCAAAATCAGAAAGAGA 1076
282 hrProValLysLysProGlyAspGlyArgLysValThrPhePheGluPro 298
|||||
1077 CTCCAGTAAAGAAACCAAGAGAGTGAAGAAAGTAAATTTTTCAGACT 1126
299 SerProGlyAspGluAsnGlyThrSerAsnLysAspGluGluPheArg 315
|||||
1127 AGCCCTGGGAGTAAAGTGAACCAAGTAAAGAGATGATGATGATGAT 1176
315 tProTyrLeuSerHisGlnGlnLeuProAlaGlyLleLeuProMetValP 332
|||||
1177 GCCCTATCTTACCATCAGACACTGCGCGGATTTCCCAATGAGTGTC 1226
332 rGluValAlaGlnAlaValAlaGlyLysGlnGlyHisThrLysAsp 348
|||||
1227 CAGAGTGTGCCAGGCTGTGCGGGTTAGTCAAGAGACACATACCAAGAT 1276
349 PheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetL 365
|||||
1277 TTGCGCAGGCGCAGCTCCAAATCTCTCCAAAGGCCAGCGTAACTGAT 1326
365 eAlaArgGluLeuLeuTyrGlyLysThrSerProThrAlaGluThrIle 382
|||||
1327 AGCCCGAAGATTTGTGTACGGGGGCACTCGCCACAGCGAGACCATTT 1376
382 eLysSerAsnLleSerSerGlyHisValProHisGlyProArgThrArg 398
|||||
1377 TAAAGATTAACATCTTTCAGACATGTACCCCATGAGACTGGCACAGA 1426
399 ProSerGluGlnLeuTyrTyrLeuSerArgAlaGlnGlyPheGlnValG 415
|||||
1427 CCCTGTGAGCAACTGTACTACTTCCAGAGCCAGGAGATTCAGAGTTGA 1476
415 uTyrLysAspPheProLysAsnAsnLysAsnGluLysValSerLeuIleA 432
|||||
1477 ATACAAAGACTTCCCAAGAACACAAAGAACGAGTGTATCTTCATCA 1526
432 snCysSerSerGlnProProLeuValSerHisGlyLleGlyLysAspVal 448
|||||
1527 ACTGCTTTCACAGCCACCTCTGTCAGGACATGGCATGGCAGAGATGTG 1576
449 GluSerCysHisAspMetAlaAlaLeuAsnLleLeuLysLleLeuSerG 465
|||||
1577 GAGTCTCTGCATGATATGCTGCACTGAACATCTTAAGGCTGCTGTGCA 1626
465 uLeuAspGlnGlnSerThrGluMetProArgThrLysAsnGlyProValS 482
1627 GTTGACCAACAGAGACAGAGATGCCAAAGGACAGGAAACGAGCAGTTT 1676
482 eAlaCysGlyArgCys 487
```

|||||  
1677 CAGCGTCGGAGCTGC 1693  
seq\_name: gb\_ro:AF227200  
seq\_documentation\_block:  
LOCUS AF227200 3059 bp mRNA  
DEFINITION Rattus norvegicus stauufen isoform Stau+16 (stau) mRNA, complete  
cds, alternatively spliced.  
ACCESSION AF227200  
VERSION AF227200.1 GI:9754865  
KEYWORDS  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 3059)  
Monshausen,M., Putz,U., Rehbein,M., Schweizer,M.,  
Desgroselliers,L., Kuhl,D., Richter,D. and Kindler,S.  
Two rat brain stauufen isoforms differentially bind RNA  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3059)  
Monshausen,M. and Kindler,S.  
Direct Submission  
TITLE Submitted (20-JAN-2000) University of Hamburg, Institute for Cell  
Biochemistry and Clinical Neurobiology, Martinistrasse 52, Hamburg  
20246, Germany  
FEATURES  
source  
1..3059  
location/Qualifiers  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/tissue\_type="brain"  
/dev\_stage="adult"  
1..3059  
/gene="Stau"  
227..1714  
/gene="Stau"  
/note="RNA-binding protein; similar to Drosophila stauufen;  
contains four RNA-binding domains; rStau+16"  
/codon\_start=1  
/product="stauufen isoform Stau+16"  
/protein\_id="AAF98119.1"  
/db\_xref="GI:9754866"  
translation="MYKPYDPHSRMOSTYSGMRGAGYAPRYFYPVPVLLYQVLS  
VGGQPNKGKRRPVKKHDATAKALRTLOSEPLPRELENGESEEENLNKSEISQV  
EIALKRLNLPVNESEFPITQVARESGPPHKNKFNVTAVSGEFGEEGSKTISKRNA  
RAVLEQLRLPLPAVERVKRIRKKSQPTCKLQTPVGGQGNPISRLAQIQAKKE  
KEPEYMLTERGLPRRREFVMQVGHHTAEGANKKVAKNAAENMLLEIGFVPO  
AOPAKPALKSEKTPVKKPGDGRKVTFFPSPDNGSNKDEFRMPLYLSHOOLPAG  
ILPMYEVAAQVAGSOGHHTKDFARAAPNKAATYATAMARELLYGGTSPMTETLS  
NISSGVPHGPTRRSEQLIYISRAQGFQVEKDKPPKNNKCEVSLINCSSOPPLVSH  
GIGKDVESCHDMAALNLIKLSLDOOSTEMPTGNGPVSACGRC"  
polyA\_signal  
3043..3048  
/gene="Stau"  
BASE COUNT 817 a 778 c 749 g 713 t 2 others  
ORIGIN  
alignment\_scores:  
Quality: 2499.00 Length: 495  
Ratio: 5.142 Gaps: 2  
Percent Similarity: 98.182 Percent Identity: 97.172  
alignment\_block:  
US-09-316-048-11 x AF227200 ..  
Align seg 1/1 to: AF227200 from: 1 to: 3059  
1 MetTyrIysProValAspProHisSerArgMetGlnSerThrTyrSery 17  
|||||  
227 ATGTATAAAGCCCGTGAGCCCACTCTCGATGCAGTCCACCTACAGCTA 276

17 rGlyMetArgGlyValAlaTyrProProArgTyrPheTyrProPhePro 34  
|||||  
277 TGGCATGCGTGGAGTGCGCTATCCCGCCAGATACTTTTACCATTTCAG 326  
34 aLProLeuLeuTyrGlnValGluLeuSerValGlyGlyGlnPhe 50  
|||||  
327 TTCACCGCTTACTCTACCAAGTTGAGCTCTCCGGAGGACAGCAGCTTT 376  
51 AsnGlyIysGlyMetArgProProValIshHisAspAlaProAlaAr 67  
|||||  
377 AATGGAGGAGAAAGATGAGACCCCTGTGAACATGATGCTACTGCGCG 426  
67 gAlaLeuArgThrLeuGlnSerGluProLeuProGluArgLeuGluVal 84  
|||||  
427 AGCCCTGAGACTGTCAGACTGAGCCCTTCCAGAGAGCTTACAGGTAA 476  
84 snGlyArgGluAlaGluGluLeuAsnLeuAsnIysSerGluIleSerGln 100  
|||||  
477 ATGGAAGAGATTCAGAAAGAAAACCTTATAATCAGAAATGAAGCCAA 526  
101 ValPheGluIleAlaLeuLysArgAsnLeuProValAsnPheGlu.... 115  
|||||  
527 GTGTTTGAGATGGCGTGAAGCGGAATTTGCTGTGAATTTTGAGCTTT 576  
116 .....ValAlaArgGluSerGlyProProHisMetLysAsn 128  
|||||  
577 CCCCTGACACAGGTGGCCCGGAGAGTGGCCGCCACATGAAGAACT 626  
128 heValThrArgValSerValGlyLeuPheValGlyGlyGlyLys 144  
|||||  
627 TTGTCACAGAGGTTTCAGTTGGGAGATTGTAGGGGAGAGAGGAGAAA 676  
145 SerLysIysIleSerLysAsnAlaAlaArgAlaValLeuGlnLe 161  
|||||  
677 AGCAAGAAGATCTCCAGAAAGATGCGCAAGGCGCTGCGGAGAGCT 726  
161 uArgArgLeuProProLeuProAlaValAlaGluArgValLysProArgIle 178  
|||||  
727 AAGGAGGCTGCGACCTTACTCTGCTGTGAGAGCGAGTCAAGCCGAGATCA 776  
178 yslLysLysSerGlnProThrCysLys.....ThrAlaProAspTyrGly 192  
|||||  
777 AGAAGAAAGATCAGCCACCTGCAAGCTTACAGACTGCCCGGACTATGAGC 826  
193 GlnGlyMetAsnProIleSerArgLeuAlaGlnIleGlnGlnAlaLys 209  
|||||  
827 CAAAGGATGATCTTATTCAGACAGCTGCGACAGATCCAGAGGCAAAA 876  
209 sGluLysGluProGluTyrMetLeuLeuThrGluArgGlyLeuProArg 226  
|||||  
877 GGAGAAAGACCGGAGTACATGCTCTTACAGAACGAGGCTTCCAGCTC 926  
226 rGArgGluPheValMetGlnValLysValGlyHisIshIshAlaGluGly 242  
|||||  
927 GCAAGGAGATTGTGATCAGGTGAGGCTTGGGCTACACCGCGAAGGA 976  
243 ValGlyThrAsnLysLysValAlaLysArgAsnAlaAlaGluAsnMet 259  
|||||  
977 GCAGGTACCAATTAAGAGGTGGCCAGCGCATGCTCTGTGAACATGCT 1026  
259 uGluIleLeuGlyPheLysValProGlnAlaGlnProAlaLysProAla 276  
|||||  
1027 GGAGATCTGCGGTTCAAAGTCCCGCCAGCGCCACCTGCCAAGCCAGCAC 1076  
276 euLysSerGluGluLysThrProValLysLysProGlyAspGlyArgLys 292  
|||||  
1077 TCAATATCAGAAAGAGAGACTCCAGTAAAGAAACAGAGAGTGAAGGAAA 1126  
293 ValThrPhePheGluProSerProGlyAspGluAsnGlyThrSerAsn 309  
|||||  
1127 GTAAATTTTTTGAACCTAGCCCTGGGGATGAAGAAATGAACCACTAATA 1176

```

309 saspgluglupheargweprottyrleuserhisglnleuproalaag 326
      ++++++|||||
1177 AGAGATGAGTTCAGATGCCCTATCTTACCATCAACAGCTGCACCGC 1226
      |||||
326 lylleuprometvalprogluvalalacinalavaliglyvalsergln 342
      |||||
1227 GGATTTCCCCCATGGTGGCCAGAGGTTGCCAGGCTGCGGGGTTAGTCAA 1276
      |||||
343 GLHSHSHSHrlyaspphethrargalaalaproasnproalalsal 359
      |||||
1277 GGACACCATACCAAGATTTCGCCAGAGGCGACTCCAAATCTCTCCAGAGC 1326
      |||||
359 arhrvalthraltmetllealaargleuleutyrglygltyserp 376
      |||||
1327 CACGGTAACAGCCATGATAGCCGAGAAATTGTGTACGGGGCACCTCGC 1376
      |||||
376 rothralagluthrilleuulysserasnllesergerglyhisvalpro 392
      |||||
1377 CCACAGCCGAGACCATTTTAAAGATACATCTCTCAGAGCATGTACCC 1426
      |||||
393 HisglYproargthrargprosergluInleuTyrtYrleuserargal 409
      |||||
1427 CATGAGACTTCGCAAGACCCCTCGAGCAACTGTACTACTCTTCAGAGC 1476
      |||||
409 agInglyPheglInvalgluTyrlYsAsppheProLYsAsnAsnLYsAsnG 426
      |||||
1477 CCAGGATTCAGGTGATATACAAAGACTTCCCAACACACAGACAGAACG 1526
      |||||
426 lucysvalserleulleasnCysserSerGlnProProleuvalSerhis 442
      |||||
1527 AGTGTATCTCTCATCACTGCTCTTCACAGCCACTCTCGTCAGGCAT 1576
      |||||
443 GLYtlleGlylsAspvalgluSerCysHisAspmetlaalaleuasnII 459
      |||||
1577 GGATTCGGCAGAGATGTGAGTCCCTGTCAATGATATGGCTGCACGAAACAT 1626
      |||||
459 eleuYlsleuSerGluLeuaspGlnGlnSerThrGluMetProargT 476
      |||||
1627 CTTAAAGCTCTGCTGAGTTGGACCAACAGACAGACAGATGCCAAGGA 1676
      |||||
476 hrGlyAsnGlyProvalSerAlaCysGlyArgCys 487
      |||||
1677 CAGGAACGAGACAGTTTCAGCGTGGCGAGGTGC 1711
      |||||
seq_name: gb_p13:AF061940

seq_documentation_block:
LOCUS AF061940 3142 bp mRNA PRI 04-MAR-1999
DEFINITION Homo sapiens staufen protein (STAU) mRNA, alternatively spliced,
complete cds.
ACCESSION AF061940
VERSION AF061940.1 GI:4335950
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 3142)
Wicham,L., Duchaine,T., Luo,M., Nabi,I.R. and DesGroselliers,L.
Mammalian staufen is a double-stranded RNA and tubulin binding
protein which localizes to the rough endoplasmic reticulum
Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
JOURNAL
MEDLINE 99147057
REFERENCE 2 (bases 1 to 3142)
DesGroselliers,L., Wicham,L. and Luo,M.
Direct Submission
JOURNAL Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900
Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada
FEATURES
source
1..3142
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"

```

```

gene /map="20q13.1"
      1..3142
      /gene="STAU"
CDS 288..1778
      /gene="STAU"
      /function="RNA binding protein"
      /note="alternatively spliced transcript"
      /codon_start=1
      /product="staufen protein"
      /protein_id="ABD17532.1"
      /db_xref="GI:4335951"
      /translation="MKLKKPKYKVPDYSRMSTYYNNMKGATPPRYEYPPYPL
LYQVELSVGGQDFNGKRTROAKHDAAKALRIQNEPLPERLEVNGRESEENLNK
SEISQVEFIALKRLNPVNEVARESEGPMMKKEFTVSGYGEVGESEKSKISKNA
ATAVEELKLEPLPAVERVKPRIKKKTRPIYKPTQSPYGGGINPISRLAIOQAKK
EKPEYTLITERGLPRRRFVMOVGVGNTAGCTGNKKVAKRNAENMLITLGFVP
OROPTPKALSKERTPIKKPKGKRVYTFPERSGPDNGISNKEDEFRMAYLSHQOLPA
GILPMWPEVAQAVVSQGHHTKDFTRAAVNPAAKAVTAMTARELILGGTSPYAEILK
NNISSGHPHGLTRPSQDLDSLRSVQGVQVEYKDFPKNNKKEFVSLNCSQPLIS
HGIGDVDSCHDMAALNIIKLKLELDQOSTEMPRTRNGPMVSCGR"

BASE COUNT 862 a 778 c 733 g 769 t
ORIGIN

alignment_scores:
      Quality: 2332.00 Length: 489
      Ratio: 4.930 Gaps: 1
Percent Similarity: 96.728 Percent Identity: 91.002

alignment_block:
US-09-316-048-11 x AF061940 ..

Align seg 1/1 to: AF061940 from: 1 to: 3142
1 MetTyrlYsProvalAspProHisSerArgMetGlnSerThrTySerTy 17
|||||
309 ATGATATAAGCTGTGTGACCCCTTACTCTCGATCGACGACCTATACTA 358
17 rGlyMetArgGlyGlyAlaTyrProProArgTyRPhetyrProPhePro 34
|||||
359 CACATATAGAGAGAGCTCTTATCCCGAGGTACTTTTATCCCATTTCCAG 408
34 alProProleuLeuTyrlGlnvalgluLeuSerValGlyGlyGlnGlnPhe 50
|||||
409 TTCACCTTTTACTTATCAAGTGAACCTTCTGTGGAGGACACACAAATT 458
51 AsnGlyLySgLyLysMetArgProProValLysHisAspAlaProAlaAr 67
|||||
459 AATGGCAAGAGAAAGACAGACAGCTGCGCAACACAGCATGCTGCTGCAA 508
67 gaLaLeuArGThrleuGlnSerGluProleuProgluArgleuGluValA 84
|||||
509 AGCGTTGAGGATCCTCGAATGAGCCCTGCCAGAGAGCTGAGAGTGA 558
84 snGlyArgGluAlaGluGluGluAsnLeuAsnLysSerGluTIIeserGln 100
|||||
559 ATGGAAGAGAAATCCGAGACAGAAAAATCTCATATAATCTGAATAAGTCAA 608
101 ValPhegluIleAlaLeuLysArgAsnLeuProValAsnPhegluValAl 117
|||||
609 GTGTTTGAGATTGCACCTTAACGGAACCTTGCTGTAATTTGAGAGTGGC 658
117 aArgGluSerGlyProProHisMetLysAsnPhelValThrArgValSerV 134
|||||
659 CCGGAGAGTGGCCACCCACATGAAGAACTTTGTGCAAGGTTTCGG 708
134 alGlygluPheValGlyGluGlyGlyLysSerTyLysIIseserLys 150
|||||
709 TTGGGAGATTGTGGGGAAGGTGAAGGAAACACAAACAAATTTCAAG 758
151 LysAsnAlaAlaArgAlaValleuGluGlnLeuArgArgLeuProProLe 167
|||||
759 AAAATATGCCCATAGCTGTCTTGAGAGAGCTGAAGAAGTTACCGCCCT 808

```

```

167 uproAlaValGluArgValIleuProArgIleuLysLysSerGluProT 184
184 hxCysLys.....ThrAlaProAspTyrGlyGlnGlyMetAsnProIle 198
859 TAGTCAAGCCACAGACAGCCCAAGATATGGCCAGGAGATCAATCCGATT 908
199 SerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTyr 215
909 ACCCGCACTGGCCAGATCCAGCGCAAGAAAAAGAGAGAGAGCAAGTA 958
215 MetLeuLeuThrGluArgGlyLeuProArgArgGluPheValMetG 232
959 CACGCTCTCCACAGACGAGCGCTCCCGCGCCGACAGGACTTTGTATGC 1008
232 LysValLysValGlyHisHisThrAlaGluGlyValGlyThrAsnLys 248
1009 AGGTGAAGGTTGGAAACCACTGCAGAAAGAACGGGCAACCAAGAAAG 1058
249 ValAlaLysArgAsnAlaAlaGluAsnMetLeuGlnIleLeuGlyPhe 265
1059 GGGCCCAAGCCGCAATGCACCGCAACATGCTGGAGATCTTGGTTTCAA 1108
265 ValProGlnAlaGlnProAlaLysProAlaLeuLysSerGluGluLysT 282
1109 AGTCCCGCAGCGGCGAGCCCAACCAACCCGACCTCAAGTCAAGAGAA 1158
282 hProValLysLysProGlyAspGlyArgLysValThrPhePheGluPro 298
1159 CACCCATTAAGAAACCAAGGAGATGAAGAAACTTATTAACCTTTTGAAC 1208
299 SerProGlyAspGluAsnGlyThrSerAsnLysAspGluGluPheArg 315
1209 GGTCTGGGGATGAATAAGGACTAGTAATAAGAGATGAGTTCAAGAT 1258
315 LProTyrLeuSerHisGlnGlnLeuProAlaGlyLeuProMetValP 332
1259 GCTTTATTCATAGTCATCAGCAGCTGCTGCTGGAATTTCTCCATGTGC 1308
332 rGluValAlaGlnAlaValGlyValSerGlnGlyHisHisThrLysAsp 348
1309 CCGAGGTGCCCGAGCTGATGAGATTAGTCAAGACATCAGACCAAGAT 1358
349 PheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMet 365
1359 TTTACACAGGCGACGCTCGAATCTGCCAAGGCCACGTAATCTCCATG 1408
365 eAlaArgGluLeuLeuTyrGlyThrSerProThrAlaGluThrIleL 382
1409 AGCCGAGAGTTGTTGTATGGGGGACCTCGCCACAGCCGAGACCATTT 1458
382 eulysSerAsnIleSerSerGlyHisValProHisGlyProArgThrArg 398
1459 TAAAGATATACATCTCTTCAGGCGACGTACCCATGACCTTCACAGAG 1508
399 ProSerGlnGlnLeuTyrTyrLeuSerArgAlaGlnGlyPheGlnVal 415
1509 CCGCTGAGCAACTGACATATCTTTCAGAGTCCAGGATTCACAGTTGA 1558
415 uTyrLysAspPheProLysAsnAsnLysAsnGlyCysValSerLeuIle 432
1559 ATACAAAGACTTCCCAAAACACAAAGAAAGATTTGTATCTCTTATCA 1608
432 sncysSerSerGlnProProLeuValSerHisGlyIleGlyLysAspVal 448
1609 ATTGCCTCTTCAGCCACTCTGATGAGCCATGATCGGCAAGAGATGTG 1658
449 GluSerCysHisAspMetAlaAlaLeuAsnIleLeuLysLeuLeuSer 465
1659 GAGTCTGCGCATGATGATGGCTGGGCGAAGCATCTTAAGTTGCTGTCTGA 1708

```

```

465 uLeuAspGlnGlnSerThrGluMetProArgThrGlyAsnGlyProValS 482
1709 GTTGGACCAACAAGATACAGAGATGCGCAAGAACAGAAACGACCATGT 1758
482 eAlaAcysGlyArgCys 487
1759 CTGTGTGTGGAGAGTGC 1775

seq_name: gb_pf3:AF061938

seq_documentation_block:
LOCUS AF061938 3217 bp mRNA PRI 04-MAR-1999
DEFINITION Homo sapiens staufen protein (STAU) mRNA, alternatively spliced,
complete cds.
ACCESSION AF061938
VERSION AF061938.1 GI:4335946
KEYWORDS
SOURCE
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 3217)
Wickham,L., Duchaine,T., Luo,M., Nabl,I.R. and Desgroselliers,L.
Mammalian staufen is a double-stranded RNA and tubulin binding
protein which localizes to the rough endoplasmic reticulum
Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
99147057
JOURNAL
MEDLINE
2 (bases 1 to 3217)
Desgroselliers,L., Wickham,L. and Luo,M.
Direct Submision
AUTHORS
TITLE
JOURNAL
Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada
FEATURES
source
1..3217
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.1"
1..3217
/gene="STAU"
363..1853
/gene="STAU"
/function="RNA binding protein"
/note="alternatively spliced transcript"
/codon_start=1
/product="staufen protein"
/protein_id="AADI7530.1"
/db_xref="GI:4335947"
/translation="MKLGKKPMYKPVDPYSMSQSTNYNMRGAYPRRYFPYVPL
LYVELSVGGQGFNGKGTROAKHDAALRLILQNEPLPERLEVNGSESEELNK
SEISOVEFIALKRLPVNFEVARSQPMKNTFKVSGVEGEGSKSKSKNA
AIALDELKRIPLPAYRKYKPKIKKTKPIYKPOTSPEVCGINPSRLAIOIAK
EKPEYTLTLERGLPRRREYVQYKVNHTAEGTGNKKVAKRNAEMLEITLFFKP
QROPTKPAKLSSEKTPIKKPGDGKKVTFEEFGSDENQTSKKEDEFRMPVLSHOOLA
GILPMVPEVDAQVNSOGHHTKDTFRAAPNPAKATVAMIARELLYGTSTPATILK
NINISGHVHPGLTRPSEOLDYLSRVQGFVYEDFPKNNKNEVSLINCSOPPLIS
HGIGKDVESCHDMAALNLIKLSLSDQSTMPRTGPMSPVCRC"

BASE COUNT 889 a 796 c 742 g 790 t
ORIGIN

alignment_scores:
Quality: 2332.00 Length: 489
Ratio: 4.930 Gaps: 1
Percent Similarity: 96.728 Percent Identity: 91.002

alignment_block:
US-09-316-048-11 x AF061938 ..
Align seg 1/1 to: AF061938 from: 1 to: 3217
1 MetTyrLysProValAspProHisSerArgMetGlnSerThrTyrSerTyr 17
|||||

```

384 ATGTATTAAGCCTGTGACCTTAAGCTGAGATGACATCACTATACTA 433  
 17 TGLYMEATGAGTGLYATATyrProPArgTyrPheTyrProPhePro 34  
 434 CAACATGAGAGGAGTGTATATCCCCGAGGACTTTTACCATTTCCAG 483  
 34 aLProPLeuLeuTyrGlnValGluLeuSerValGlyGlyGlnGlnPhe 50  
 484 TTCCACCTTTACTTATCAAGTGAAGTCTTCTGGGAGGACAGCAATTT 533  
 51 AsnGlyLysGlyLysMetArgProProValLysHisAspAlaArg 67  
 534 AATGGCAAGGAAAGACAGACAGGCTCGCAAAACACGATGCTGCTGCAA 583  
 67 gAlaLeuArgThrLeuGlnSerGluProLeuProGluArgLeuGlnVal 84  
 584 AGCTTTGAGATCTCTGAGATGAGCCCTGCGCAGAGAGCTGGAGGTGA 633  
 84 snGlyArgGluAlaGluGluGlnLeuAsnLeuAsnLysSerGluLysSerGln 100  
 634 ATGGAAGAGATCGCAAGAGAAATCTCAATAATCTGAATAATGACAA 683  
 101 ValPheGluValLeuAlaLeuLysArgAsnLeuProValAsnPheGluValAl 117  
 684 GTGTTTGAGATTGCACTTAAACGAACTGCTGTGATTTCCAGGTGGC 733  
 117 ArgGluSerGlyProProHisMetLysAsnPheValThrArgValSerV 134  
 734 CCGGAGAGAGTGGCCACCCACATGAAAGAACTTTGTACCAAGGTTTCGG 783  
 134 aLgLyLysPheValGlyGlyGlyLysSerLysLysLysLysLys 150  
 784 TTGGGAGGTTTGTGGGGAAGTGAAGGAAAGCAAGAAATTTTCAAG 833  
 151 LysAsnAlaAlaArgAlaValLeuGluGlnLeuArgArgLeuProPole 167  
 834 AAAATGCTCCCATAGCTTCTTGAAGAGCTGTAAGAACTTACCGCCCT 883  
 167 uProAlaValGluArgValLysProArgTyleLysLysLysSerGlnProT 184  
 884 GCCGCAAGTGAACAGTAAGCTTAAGATCAAAAAGAAACAAACCA 933  
 184 hrcLysLys.....ThrAlaProAspTyrGlyGlnGlyMetAsnProIle 198  
 934 TAGTCAGACCAACAGACCAAGCCCAAGATATGCGCGGAGATCAATCCGATT 983  
 199 SerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTyr 215  
 984 AGCCGACTGGCCCGATCCAGACGCAAAAAGAGAGAGCCAGATA 1033  
 215 rMetLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMetG 232  
 1034 CACGCTCTCAACAGAGCGAGGCTCCCGCGCGAGGAGTGTGTGATGC 1083  
 232 InValLysValGlyHisHisThrAlaGluGlyValGlyThrAsnLysLys 248  
 1084 AGGTGAAGGTTTGGAAACACACACTCCAGAGAAAGGCGCACCAACAGAG 1133  
 249 ValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPheLys 265  
 1134 GTGGCCCAAGGCAATGCAAGCCGAGAACATGCTGTGAGATCTCTGTTCAA 1183  
 265 sValProGluAlaGlnProAlaLysProAlaLeuLysSerGlnGluLysT 282  
 1184 AGTCCCGCAGCGGACCCACCAACCCGCACTCAAGTCAGAGAGAGAGA 1233  
 282 hrProValLysLysProGlyAspGlyArgTyrValThrPhePheGluPro 298  
 1234 CACCCATAAAGAAACAGGAGGATGAGAAAGAAATACCTTTTGAACCT 1283  
 299 SerProGlyAspGluAsnGlyThrSerAsnLysAspGluLysPheArgMe 315  
 1284 GCCTCTGGGATGAAATGGAGACTAGTAATAAAGAGATGAGTTCAAGAT 1333

315 tProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuProMetValP 332  
 1334 GCCTTATCTAAGTATCATCAGACGCTGCTGCAATCTTCCCATGGTGC 1383  
 332 roGluValAlaGlnAlaValGlyValSerGlnGlyHisHisThrLysAsp 348  
 1384 CCGAGGTGCGCCAGCTGTAGAGAGTAGTCAAGSACATCACACCAAGAT 1433  
 349 PheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetIle 365  
 1434 TTTAACAGGGGAGCTCCGAATCTGCGCAAGGCCACGGTAACTGCCATGAT 1483  
 365 eAlaArgGluLeuLeuTyrGlyThrSerProThrAlaGluThrIleL 382  
 1484 AGCCCGAGAGTGTGTATGGGGCCACCTGCGCCACAGCCGAGCCATTT 1533  
 382 euLysSerAsnLysSerSerGlyHisValProHisGlyProArgThrArg 398  
 1534 TAAAGATTAACATCTCTTCAAGGCCAGTACCCCATGAGCTCTCAGCAGA 1583  
 399 ProSerGluGlnLeuTyrTyrLeuSerArgAlaGlnGlyPheGlnValG 415  
 1584 CCTCTGAGCAACTGAGACTATCTTCCAGACTCCAGGAGATTCCAGTTGA 1633  
 415 uTyrLysAspPheProLysAsnLysAsnGluCysValSerLeuIleA 432  
 1634 ATACAAAGACTCCCAAAACACAGAACAGAAATTTGTATCTTTATCA 1683  
 432 snCysSerSerGlnProProLeuValSerHisGlyIleGlyLysAspVal 448  
 1684 ATTCCTCTCTCAGCCACCTCTGATCAGCCATGGTATCGGCAAGAGTGTG 1733  
 449 GluSerCysHisAspMetAlaAlaLeuAsnIleLeuLysLeuLeuSerG 465  
 1734 GAGTCCGTCATGATATGCGCTGCAAACTTAAAGTTGCTGTCTGA 1783  
 465 uLeuAspGlnGlnSerThrGluMetProArgThrGlyAsnGlyProValS 482  
 1784 GTTGACCAACAAAGTACAGAGATGCCAAGACAGAAAGACGACCAATGT 1833  
 482 eraLacySGlyArgCys 487  
 1834 CTGTGTGGGAGGTGC 1850  
 seq\_name: gb\_p13:AF061941  
 seq\_documentation\_block:  
 LOCUS AF061941 3349 bp mRNA PRI 04-MAR-1999  
 DEFINITION Homo sapiens staufen protein (STAU) mRNA, alternatively spliced,  
 complete cds.  
 ACCESSION AF061941  
 VERSION AF061941.1 GI:4335952  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 3349)  
 Wickham,L., Duchaine,T., Luo,M., Nabl,I.R. and DesGroselliers,L.  
 Mammalian staufen is a double-stranded RNA and tubulin binding  
 protein which localizes to the rough endoplasmic reticulum  
 Mol. Cell. Biol. 19 (3), 2220-2230 (1999)  
 JOURNAL  
 MEDLINE 99147057  
 REFERENCE 2 (bases 1 to 3349)  
 DesGroselliers,L., Wickham,L. and Luo,M.  
 AUTHORS Direct Submission  
 TITLE Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900  
 JOURNAL Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada  
 FEATURES  
 source  
 1..3349  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"

/chromosome="20"  
 /map="20q13.1"  
 1..3349  
 /gene="STAU"  
 495..1985  
 /gene="STAU"  
 /function="RNA binding protein"  
 /note="alternatively spliced transcript"  
 /codon\_start=1  
 /product="stau protein"  
 /protein\_id="AA017533.1"  
 /db\_xref="GI:4335953"  
 /translation="MKLKKPKPKVPDPYSMSTYNYNMKGAVPPREYFPFVPL  
 LYQVELSVGGQFNGKGTROAAKHDAAKALRLIIONEPLPERLEVNGRESEENLNK  
 SEISOVEIALKRLNLPVNFVEARSGPPHMKVTVKSVGEVEGSKSKISKNA  
 AIAVLEELKLPPLPAVERVKPRIKKTKPIVKPOTSPVGGGINPISRLAIOQAKK  
 EKEBEYTLLEERGLPRRREPVOYKVNHTAEIGSTNKVKAKRAEMLEITLFKYP  
 OKOPTKALNSEKTPTRKPKPDGKKVTFEEFGSDENGTSKNKEDEFRMPYLSHOOLA  
 GILMPPEVAOAVSQGHHTKDFTRAAPNPAKATVYAMIARELLGGTSTPAETLK  
 NMISGHVPPLTRPSEOLDLSRVQGFQVEYKDFPKNNKNEVSLINCSOPPLIS  
 HGICKDYESCHDMAALNLIKLSLDOOSTEMPRGNGPMSVCGRC"

BASE COUNT 913 a 838 c 777 g 821 t  
 ORIGIN

alignment\_scores:  
 Quality: 2332.00 Length: 489  
 Ratio: 4.930 Gaps: 1  
 Percent Similarity: 96.728 Percent Identity: 91.002

alignment\_block:  
 US-09-316-048-11 x AF061941 ..

Align seg 1/1 to: AF061941 from: 1 to: 3349

1 MetTyrLysProValAspProHisSerArgMetGlnSerThrTyrSerTyr 17  
 |||||  
 516 ATGTATAAGCCTGTGACCTTACTCTCGATGCAGTCCACCTATAACTA 565  
 17 rglMetLarGlyGlyAlaTyrProProArgTyrPheTyrProPhePro 34  
 |||||  
 566 CAACATGAGAGAGGAGTCTTATCCCGGAGTACTTTACCCATTTCACG 615  
 34 alProLeuLeuTyrGlnValGlnLeuSerValGlyGlnGlnPhe 50  
 |||||  
 616 TTCACCTTACTTATCAAGTGAACCTTCTGTGGAGGACACGCAATTT 665  
 51 AsnGlyLysGlyLysMetArgProProValLysHisAspAlaProAla 67  
 |||||  
 666 AATGGCAAGAGAAAGACAGACAGCTCGGAACACAGATGCTGTCGCA 715  
 67 gAlaLeuAlaGThrLeuGlnSerGluProLeuProGluArgLeuGlnVal 84  
 |||||  
 716 ACCGTTGAGAGATCTGCAAGATAGCCCTGCGCAGAGAGCTGGAGGTA 765  
 84 sngLYArgGlnAlaGlnGlnGlnLysLeuAsnLysSerGlnLysSerGln 100  
 |||||  
 766 ATGGAAAGAAATCCGAAAGAAATCTCAATTAATCTGAATAATGCTAA 815  
 101 ValPheGluLLeuAlaLeuLysArgAsnLeuProValAsnPheGluVal 117  
 |||||  
 816 GTGTTGAGATTGCACTTAACGGAACCTGCGTGAATTTGAGAGTGGC 865  
 117 aArgGlnSerGlyProProHisMetLysAsnPheValThrArgValSer 134  
 |||||  
 866 CCGGGAGAGTGGCCACCCACATGAAGACTTTTGACCAAGCTTTCGG 915  
 134 aAlGlyLupPheValGlyGlnGlyLysSerLysLysLysSerLys 150  
 |||||  
 916 TTGGGAGAGTTTGTGGGGAAGGTGAAGGAAAGCAAGAGATTTTAAAG 965  
 151 LysAsnAlaAlaArgAlaValLeuGlnGlnLeuArgArgLeuProLeu 167  
 |||||

966 AAAAATGCCGCATACCTGTCTTTCAGAGAGCTGCAAGAAGTTACCGCCCT 1015  
 167 uProAlaValGluArgValLysProArgIleLysLysLysSerGlnPro 184  
 |||||  
 1016 GCCTGCAGTTGAACGAGTAAGCTTAGAATCAAAAAGAAAAACCAACCA 1065  
 184 hCysLys.....ThrAlaProAspTyrGlyGlnLysMetAsnProIle 198  
 |||||  
 1066 TACTCAAGCCACAGACAGCCAGATATGCGCAGGAGGATCAATCCGATT 1115  
 199 SerArgLeuAlaGlnIleGlnGlnAlaLysGlyLysGlnProLys 215  
 |||||  
 1116 AGCGGACTGCGCCAGATCCAGCAGGCAAAAAAGCAAGAGGCAAGTA 1165  
 215 rMetLeuLeuThrGluArgGlyLeuProArgArgGluPheValMetG 232  
 |||||  
 1166 CAGCTCTCTCAGAGAGAGGCTCCCGCGCCGAGGAGATTGTATGC 1215  
 232 LysValLysValGlyHisHisThrAlaGlnGlyValGlyThrAsnLys 248  
 |||||  
 1216 AGGTGAAGGTTGGAACCACTGCAAGAAAGAGGAGGCAACCAAGAA 1265  
 249 ValAlaLysArgAsnAlaAlaGluAsnMetLeuGlnIleLeuGlyPhe 265  
 |||||  
 1266 GTGGCCAGCGCAATGCAGCCGGAACATGCTGAGATCCTTGCTTCAA 1315  
 265 sValProGlnAlaGlnProAlaLysProAlaLysSerGlnGluLys 282  
 |||||  
 1316 AGTCCCGCAGCGGACGCCACCAACCCGCTCAAGTCAAGAGAAAGAA 1365  
 282 hProValLysLysProGlyAspGlyArgLysValThrPhePheGluPro 298  
 |||||  
 1366 CACCCTAAAGAAACCAAGGAGATGAAGAAAGTAACTTTTGAACCT 1415  
 299 SerProGlyAspGluAsnGlyThrSerAsnLysAspGluGluPheArg 315  
 |||||  
 1416 GGCTCTGGGAGATGAAATGGAATGACTAGTAATAAAGAGATGAGTTCA 1465  
 315 tProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuProMetVal 332  
 |||||  
 1466 GCCTTATCTAAGTCATCAGACGCTGCTGGAAATCTTCCATGGTGC 1515  
 332 rGluValAlaGlnAlaValAlaGlyValSerGlnGlyHisThrLysAsp 348  
 |||||  
 1516 CCGAGCTCCGCCAGGCTGTAGAGAGTTAGTCMAAGCAATCAACCAAGAT 1565  
 349 PheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMet 365  
 |||||  
 1566 TTTACAGGGCAGCTCGAATCTGCGCAAGGCCAGCGTAATCCATGAT 1615  
 365 eAlaArgGluLeuLeuTyrGlyThrSerProThrAlaGluThrIle 382  
 |||||  
 1616 AGCCGAGAGTTGTTATGAGGGGACCTGCGCCACAGCGCAACCAATTT 1665  
 382 eLysSerAsnLysSerSerGlyHisValProHisGlyProArgThrArg 398  
 |||||  
 1666 TAAAGATAAATCAATCTTCAAGCCAGCTACCCATGAGACTCTCAGAGA 1715  
 1716 CCCTCTGACCAACTGGACTATCTTTCAGAGTCCAGAGGATTCAGAGTTGA 1765  
 415 uTyrLysAspPheProLysAsnAsnLysAsnGlyCysValSerLeuIle 432  
 |||||  
 1766 ATACAAGACTTCCCAAAAACAACAAGAAAGCAATTTGTAATCTTATACA 1815  
 432 sNcysSerSerGlnProProLeuValSerHisGlyIleGlyLysAspVal 448  
 |||||  
 1816 ATTGCTCTCTCAGCCACTCTGATCAGCAATGATTCGGCAAGATGAG 1865  
 449 GluSerCysHisAspMetAlaAlaLeuAsnLLeuLysLeuLeuSer 465  
 |||||  
 1866 GAGTCTCGCATGATGATGCTGCGTGAACATCTTAAAGTTGCTGTCTGA 1915



```

465 uleuaspGLInserThrcIumetProArGThrGlyAsnGlyProValS 482
|||||
1916 GTTGGACCAACAAAGTACAGATGCCAAGACGAGAACGACCAATGT 1965
482 eraIaCysGlyArgCys 487
|||||
1966 CTGTGTGTGGAGGTGC 1982

seq_name: gb_p13:AF061939

seq_documentation_block:
LOCUS AF061939 3506 bp mRNA PRI 04-MAR-1999
DEFINITION Homo sapiens staufen protein (STAU) mRNA, alternatively spliced,
complete cds.
ACCESSION AF061939
VERSION AF061939.1 GI:4335948
KEYWORDS
SOURCE human.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 3506)
Wickham,L., Duchaine,T., Luo,M., Nabil,I.R. and Desrosiers,L.
Mammalian staufen is a double-stranded RNA and tubulin binding
protein which localizes to the rough endoplasmic reticulum
Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
JOURNAL
99147057
MEDLINE
2 (bases 1 to 3506)
Desrosiers,L., Wickham,L. and Luo,M.
Direct Submission
Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900
Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada
FEATURES
Source Location/Qualifiers
1..3506
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.1"
1..3506
/gene="STAU"
409..2142
/gene="STAU"
/function="RNA binding protein"
/note="alternatively spliced transcript"
/codon_start=1
/product="staufen protein"
/protein_id="AAD17531.1"
/db_xref="GI:4335949"
/translation="MSOVQVONPSAALSGSOLINKNOSLISOPLMISPTSSLPS
EAGRPIONSALPSAISTSTSAASITPTVFLNLMCKLGKKPKPKPDVPSRMST
YNNMRGATPPRYFPYPPVPLLYOVELSVGGQFNGCKTQAKHDAKHAALKEIL
ONEPLPERLEVNGRESEENLNKSEISQVETALRNKLVNEVAKESGPPMKNFVT
KVSVEFVGESEKSKISKNAALAVLEELKPLPAVEVKKPIKKTKPIVKPO
TSPEVGGINPISRLAQIOAKKEKEPEYTLLEGLPRREVMQVKGNHTABGTG
TKKAKRNAENMLIELGFKVPOPOPTPALSKSEKPIKKGQGRKVTPEPGSGD
EMGTNKKDEFEFMYLISHOOLPAGILPMPEVAOVAVGSGHHTDFPTAAANPKAT
VTAMTARELLYIGTSPATETILKNNISSGHVPHGLTPRSEDLTISRQGVQVEKQ
PFKNKNEFVSLINCSOPPLISHGIGKDVESCHDMAALNIIKLISELDQOSTEMP
TGMPMSVCGRC"
BASE COUNT 965 a 875 c 790 g 876 t
ORIGIN
alignment_scores:
Quality: 2332.00 Length: 489
Ratio: 4.930 Gaps: 1
Percent Similarity: 96.728 Percent Identity: 91.002

alignment_block:
US-09-316-048-11 x AF061939 ..
Align seg 1/1 to: AF061939 from: 1 to: 3506

```

```

1 MetTyrLysProValAspProHisSerArgMetGlnSerThrTyrSerTy 17
|||||
673 ATGTATTAAGCTCTGTGACCTTACTCTCGGATGACATGCCACTTAAC 722
17 rGlyMetArgGlyGlyAlaTyrProProArgTyrPheTyrProPhePro 34
|||||
723 CAACATGAGAGAGAGAGCTTATCCCGGAGTACTTTTACCATTTCAG 772
34 aLProProLeuLeuTyrGlnValGluLeuSerValGlyGlyGlnGlnPhe 50
|||||
773 TTCACCTTTACTTATCAAGTGAACCTTCTGTGGAGGACACCAATTT 822
51 AsnGlyLysGlyLysMetArgProProValIshHisAspAlaProAla 67
|||||
823 AATGCCAAGGAAGACACAGACAGCTGCCGAAACACATGCTGCTGCCAA 872
67 gAlaLeuArgThrLeuGlnSerGluProLeuProGluArgLeuValAla 84
|||||
873 AGCGTTGAGGATCCTGCAGAAATGAGCCCTGCCAGAGAGCTGAGGTGA 922
84 snGlyArgGluAlaGluGluGlnLeuAsnLysSerGluIleSerGln 100
|||||
923 ATGGAAGAGAAATCCGAAGAGAAATCTCAATTAATCTGAATTAAGTCAA 972
101 ValPheGluIleAlaLeuLysArgAsnLeuProValAsnPheGluValAl 117
|||||
973 GTGTTGAGATGACATCAATTAAGGAACTTGCTGTGAATTCGAGGTGCC 1022
117 aArgGluSerGlyProProHisMetLysAsnPheValThrArgValSerV 134
|||||
1023 CCGGAGAGAGTGGCCACCCACATGAAGAACTTGTGACCAAGGTTTCGG 1072
134 aLgLyGluPheValGlyGluGluGlyLysSerLysLysIleSerLys 150
|||||
1073 TTGGGAGGTTTGTGGGGAAGGTGAAGGAAACCAAGCAATTTTCAAG 1122
151 LysAsnAlaAlaArgAlaValLeuGluGlnLeuArgArgLeuProPole 167
|||||
1123 AAAAATGCCCGCATAGCTGTTCTTGAGAGAGTGAAGAAATTCACGCCCT 1172
167 uProAlaValGluArgValLysProArgIleLysLysLysSerLysProT 184
|||||
1173 GCCTGCAAGTTGAACGAGTAAGCCTAGAAATCAAAAAGAAACCAACCCA 1222
184 hrCysLys.....ThrAlaProAspTyrGlyGlnGlyMetAsnProIle 198
|||||
1223 TAGTCAAGCCACAGACAAAGCCAGAAATATGCCCAGGGGATCAATCCGATT 1272
199 SerArgLeuAlaGlnIleGlnGlnAlaAlaLysLysGluLysGluProGluTy 215
|||||
1273 AGCGGACTGCGCCAGATCCACGACGCAAAAAGAGAGAGCCAGAGTA 1322
215 tMetLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMetG 232
|||||
1323 CACCTCTCCTCACAGAGGAGCCTCCGCGCGCAGGAGGATTTGTGATGC 1372
232 tValLysValGlyIshHisThrAlaGluGlyValGlyThrAsnLysLys 248
|||||
1373 AGGGAAGGTTTGAACACCACTGACAGAAAGACGGGACCAACAAGAG 1422
249 ValAlaLysArgAsnAlaAlaGlnAsnMetLeuGluIleLeuGlyPheLy 265
|||||
1423 GTGGCCAAAGCCCAATGCAGCCGAGACATGCTGGAGATCCTTGGTTTCAA 1472
265 sValProGlnAlaGlnProAlaLysProAlaLeuLysSerGluGluTy 282
|||||
1473 AGTCCCGAGCGGAGCCACCAACCGCAGCTCAAGTCAAGAGAGAGAGA 1522
282 hrProValLysLysProGluLysPglYArgLysValThrPhePheGluPro 298
|||||
1523 CACCATATAAGAAACAGAGGGAGTGAAGAAAGTAACTTTTGTGAACCT 1572

```

299 SerProGlyAspGluAsnGlyThrSerAsnLysAspGluGluPheArgGme 315  
::: |||||||||||||||||||||||||||||||||||||||||||||||||||  
1573 GGCTGTGGGATGAAATGAGTACTAGTAATTAAGAGATGATTCAGAT 1622  
315 tProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuPrometValP 332  
||||| |||||||||||||||||||||||||||||||||||||||||||||||  
1623 GCGTTATCTAAGTCATCAGCAGCTGCTCGTGAATCTTCCATGCTGCG 1672  
332 roGluValAlaGlnAlaValAlaGlyValSerGlnGlnHisThrLysAsp 348  
||||| |||||||||||||||||||||||||||||||||||||||||||||||  
1673 CCGAGTCCGCCAGGCTGTAGAGGTTAGTCAGGACATCAGACCAAGAT 1722  
349 PheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetIi 365  
||||| |||||||||||||||||||||||||||||||||||||||||||||||  
1723 TTATACAGGACGACCTCCGAATCTCTGCCAAGGCCAGGTAACTGCATGAT 1772  
365 eAlaArgGluLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleL 382  
||||| |||||||||||||||||||||||||||||||||||||||||||||||  
1773 AGCCCGAGAGTTGTTGTATGGGGGACCTCGCCACAGCCGAGACCATTT 1822  
382 euLysSerAsnLysSerSerGlyHisValProHisGlyProArgThrArg 398  
||||| |||||||||||||||||||||||||||||||||||||||||||||||  
1823 TAAAGAAATACATCTTCTTCAGGCCACGTACCCATGACCTCTCAGAGAG 1872  
399 ProSerGluGlnLeuTyrTyrLeuSerArgAlaGlnGlyPheGlnValG 415  
||||| |||||||||||||||||||||||||||||||||||||||||||||||  
1873 CCTCTGAGCAACTGGACTATCTTTCAGAGTTCAGAGATTCAGAGTTGA 1922  
415 uTyrLysAspPheProLysAsnLysAsnGlyCysValSerLeuIleA 432  
||||| |||||||||||||||||||||||||||||||||||||||||||||||  
1923 ATACAAAGACTTCCCAAAACAAAGAAAGCAATTTGTATCTTCTATCA 1972  
432 sncCysSerSerGlnProProLeuValSerHisGlyIleGlyLysAspVal 448  
||||| |||||||||||||||||||||||||||||||||||||||||||||||  
1973 ATTGCTCTCTCAGCCACCTCTGATCAGCATGTATCGCAAGAGATG 2022  
449 GluSerCysHisAspMetAlaAlaLeuAsnIleLeuLysLeuLeuSerG 465  
||||| |||||||||||||||||||||||||||||||||||||||||||||||  
2023 GAGTCTGCGCATGATATGGCTGCGCTGAACATCTTAAAGTGTGCTGA 2072  
465 uLeuAspGlnGlnSerThrGluMetProArgThrGlyAsnGlyProValS 482  
||||| |||||||||||||||||||||||||||||||||||||||||||||||  
2073 GTTGGACCAACAAGTACAGAGATGCCAAGAACAGGACCAATGT 2122  
482 eAlaCysGlyArgCys 487  
||::: |||||||||||||||  
2123 CTGTGTGTGGAGGTGC 2139  
seq\_name: gb\_pr4:AK025519  
seq\_documentation\_block:  
LOCUS AK025519 3194 bp mRNA 29-SEP-2000  
DEFINITION Homo sapiens cDNA: FLJ21866 fls, clone HEP02379, highly similar to  
AF061938 Homo sapiens staufen protein (STAU) mRNA.  
ACCESSION AK025519  
VERSION AK025519.1 GI:10438059  
KEYWORDS oligo capping; fls (full insert sequence).  
SOURCE Homo sapiens hepatoma cell\_line:Hepg2 cDNA to mRNA, clone\_11b:HEP  
clone:HEP02379.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (sites)  
REFERENCE  
AUTHORS Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,  
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,  
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.  
NEO human cDNA sequencing project  
Unpublished (2000)  
2 (bases 1 to 3194)  
Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,  
Shibahara,T., Tanaka,T. and Nakamura,Y.  
TITLE JOURNAL  
REFERENCE  
AUTHORS  
TITLE Direct Submission  
JOURNAL Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio

Sugano, Institute of Medical Science, University of Tokyo,  
Laboratory of Genome Structure Analysis, Human Genome Center;  
Shirokane-dai, 4-6-1, Minato-Ku, Tokyo 108-8639, Japan  
(E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,  
Fax:81-3-5449-5416)  
COMMENT  
NEO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing: Research Association for Biotechnology; cDNA library  
construction, 5'- & 3'-end one pass sequencing: Department of  
Virology and Human Genome Center, Institute of Medical Science,  
University of Tokyo (partly supported by Science and Technology  
Agency).  
FEATURES  
source  
1. 3194  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="hepg2"  
/cell\_type="hepatoma"  
/clone="HEP02379"  
/clone\_11b="HEP"  
/note="cloning vector pME18SFL3"  
misc\_feature  
1..3194  
/note="highly similar to AF061938 Homo sapiens staufen  
protein (STAU) mRNA"  
BASE COUNT 894 a 787 c 732 g 781 t  
ORIGIN  
alignment\_scores:  
Quality: 2275.50 Length: 489  
Ratio: 4.801 Gaps: 2  
Percent Similarity: 96.933 Percent Identity: 90.593  
alignment\_block:  
US-09-316-048-11 x AK025519 ..  
Align seg 1/1 to: AK025519 from: 1 to: 3194  
1 MetTyrLysProValAspProHisSerArgMetGlnSerThrTyrSerTy 17  
||||| |||||||||||||||||||||||||||||||||||||||||||||||  
353 ATGTATTAAGCTGTGTGACCTTACTCTCGATGACATCCACCTATTAAC 402  
17 rGlyMetArgGlyGlyAlaTyrProProArgTyrPheTyrProPhePro 34  
||||| |||||||||||||||||||||||||||||||||||||||||||||||  
403 CAACATGAGAGGAGGAGTGTATCCGCCAGAGTACTTTACCCATTTCCAG 452  
34 aLProProLeuLeuTyrGlnValAlaGluLeuSerValGlyGlyGlnGlnPhe 50  
||||| |||||||||||||||||||||||||||||||||||||||||||||||  
453 TTCCACCTTACTTATCAAGTGAAGTCTTCTGTGGAGGACAGCAATTT 502  
51 AsnGlyLysGlyLysMetArgProProValLysHisAspAlaProAlaAr 67  
||||| |||||||||||||||||||||||||||||||||||||||||||||||  
503 AATGGCAAGGAAAGACAGACAGGCTGCGAAGACGATGCTGCTGCCAA 552  
67 gAlaLeuArgThrLeuGlnSerGluProLeuProGluArgLeuGluValA 84  
||||| |||||||||||||||||||||||||||||||||||||||||||||||  
553 AGCGTTGAGAGATCTTCAGAAATGAGCCCTGCCAGAGAGCTGAGAGTGA 602  
84 sngLysArgGluAlaGluGluGluAsnLeuAsnLysSerGluLysSerGln 100  
||||| |||||||||||||||||||||||||||||||||||||||||||||||  
603 ATGGAAAGAGAAATCCGAAGAAAGAAATCTCAATAAATCTAAATTAAGTCAA 652  
101 ValPheGluIleAlaLeuLysArgAsnLeuProValAspPheGluValAl 117  
||||| |||||||||||||||||||||||||||||||||||||||||||||||  
653 GGTGTTGAAGATTGCACTTAACGGAACTTGCTGTAATTTGAGAGTGGC 702  
117 aArgGluSerGlyProProHisMetLysAsnPheValThrArgValSerTy 134  
||||| |||||||||||||||||||||||||||||||||||||||||||||||  
703 CCGGAGAGATGCGCCACCCACATGAGAATCTTGTGACCAAGGTTTGG 752  
134 aGlyGluPheValGlyGlyGlyGlyLysSerLysLysLysLys 150  
||||| |||||||||||||||||||||||||||||||||||||||||||||||  
753 TTGGGAGTTTGTGGGGGAGGTGAAGGAAAGCAAGGAAATTTCAAG 802

```

449 1uScerYsthisApmetAlaIalaleuAsnIleleuLysleuSergu 465
|||||
1702 AGTCCGCGCATGATATGGCTGCGCTGAACATCTTAAAGTGTGCTGTGAG 1751
466 LeuASpGIgInSerThrlumetProATGTrnGIYasngIYproValse 482
|||||
1752 TTGGACCAACAAAGTACAGATGCCAACAACAGGAACGACCAATGTC 1801
482 rAlAcysGIYarGcys 487
|||||
1802 TGTGTGTGGGAGGTGC 1817

seq_name: gb_p6: HSA132258

seq_documentation_block:
LOCUS HSA132258 3066 bp mRNA PRI 14-MAY-1999
DEFINITION Homo sapiens mRNA for staufen protein, partial.
ACCESSION AJ132258
VERSION AJ132258.1 GI:4572587
KEYWORDS staufen gene; staufen protein.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS Falcon,A.M., Fortes,P., Marion,R.M., Beloso,A. and Ortin,J.
TITLE Interction of influenza virus NS1 protein and the human homologue
of Staufen in vivo and in vitro
JOURNAL Nucleic Acids Res. 27 (11), 2241-2247 (1999)
MEDLINE 99263025
REFERENCE 2 (bases 1 to 3066)
AUTHORS Ortin,J.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-1999) Ortin J., Molecular and Cellular Biology,
Centro Nacional de Biotecnologia, Cantoblanco, 28049 Madrid, SPAIN

FEATURES
Source
1..3066
/organism="Homo sapiens"
/db_xref="taxon:9606"
36..1526
/gene="staufen"
36..1526
/gene="staufen"
/function="dsRNA binding protein"
/codon_start=1
/product="staufen protein"
/protein_id="CABA0082.1"
/db_xref="GI:4572588"
/translation="MKLGKKPKYKRPVDPSIMOSTYINNMKGAYPRRYPPYVPL
LYOVLVGGQGFNGKTRQAAHDAAKALRLTONEPLPERLEVNGRSESEENLNK
SEISOVEIALKRNLPVNFVARESGAPPHMKPTKSVGFEVGEGBSKSKSKNA
AIAVEELKRLPLPAYERVKPRIKTKTKPIVKFOGTSPEVGGINPISRLAIQAKK
EKKEPYLLTFRGILPBRREFYMOYKVNHTAEGCTKKVAKRNAENMLEIGFVYD
ORSQOTRQVNGEPTIKETGRKSKNSLFINALMENGTSKKEDFPMPTYSHOOLPA
GILDPVEPAOAVGVSOGHHKTDTRAPNPAKATVAMKARELLIGTSEPTAETILK
NINSGVHPGPILRPSQDLYLSRVGQFOVEYEDFKNKNNEFVSLINCSQPLPIS
HGICRDVESCDDMAALNITKLSELDOSTEMPTGNGPMVSGRC"

BASE COUNT 891 a 693 c 699 g 783 t
ORIGIN

alignment_scores:
Quality: 2274.50 Length: 492
Ratio: 4.819 Gaps: 5
Percent Similarity: 95.935 Percent Identity: 90.244

alignment_block:
US-09-316-048-11 x HSA132258
Align seg 1/1 to: HSA132258 from: 1 to: 3066

1 MetTyrLysProValAspProHisSerArgMetGlnSerThrTyrSery 17
|||||

```

```

57 ATGTATAAGCCTGTGACCTTACTCTCGGATGCGAGTCCACCTATAACTA 106
17 rglymeLarGlyGlyAlaTyrProProlArgTyrPheTyrProPhePro 34
107 CAACATGAGAGGAGGTCCTTATCCCGGAGTACTTTTACCCATTTCCAG 156
34 aLProProlLeuTyrGlnValAlaGluSerValAlaGlyGlnGlnPhe 50
157 TTCACCTTACTTATCAAGTGAAGTCTTCTGCGAGAGCAGCAATTT 206
51 AenGlyLysGlyLysMetArgProProlValLysHisAspAlaProAla 67
207 AATGCGAAGAGAAAGACAAAGACAGGCTCGCAACACAGATCTGCTCAA 256
67 gAlaLeuArgThrLeuGlnSerGluProLeuProGluArgLeuGluVal 84
257 AGCGTTGAGATCTCGAATATGAGCCCTCGCAGAGAGGCTGAGGTGA 306
84 snGlyArgGluAlaGluGluGluValAsnLeuAsnLysSerGluLysSer 100
307 ATGGAAGAGATCCGAAAGAAATCTCATTAATCTGAATAATGTCAA 356
101 ValPheGluLeuAlaLeuLysArgAsnLeuProValAsnPheGluVal 117
357 GTGTTGAGATTGCACTTAACCGAACTTGCCGTGAATTTGAGGTGGC 406
117 aArgGluSerGlyProProlHisMetLysAsnPheValThrArgValSer 134
407 CCGGGAGAGTGGCCACCCACATGAAAGACCTTGTGACCAAGGTTTCG 456
134 aLgLyLysPheValGlyGluGlyGlyLysSerLysLysLysSerLys 150
457 TTGGGAGATTGTGGGGAAGCTGAAGGGAAGCAAGAAATTCAGAG 506
151 LysAsnAlaAlaArgAlaValLeuGluGlnLeuArgArgLeuProLe 167
507 AAAAATGCGCCCATGCTTCTTGTGAGAGCTGAAGAAAGTTACGCCCT 556
167 uProAlaValAlaGluArgValLysProArgLysLysLysSerGluPro 184
557 GCCTGAGTGGAAACGAGTAAAGCTTAGAATCAAAAAGAAACAAACCCA 606
184 hrcLysLys.....ThrAlaProAspTyrGlyGlnLysMetAsnPro 198
607 TACTCAAGCCACAGACAGACGCCAGAAATATGGCCAGGAGATCCGATT 656
199 SerArgLeuAlaGlnLeuGlnAlaValLysGluLysGluProGlu 215
657 AGCCGACTGCCCCAGATCCAGCAGGCAAAAAGAGAGAGCCAGCAAGTA 706
215 rMeLeuLeuThrGluArgGlyLeuProArgArgGluPheValMetC 232
707 CACGCTCTCAGACAGCAGAGCTCCCGCCCGCCAGGAGATTTGTATGC 756
232 InValLysValGlyHisHisThrAlaGluGlyValGlyThrAsnLys 248
757 AGGTGAAGGTTGGAAACCACTGCAAGAGAGAGGCAACCAACAAGAG 806
249 ValAlaLysArgAsnAlaAlaGluAsnMetLeuGluLeuLeuGlyPhe 255
807 GTGGCCCAACCGCAATGACCCGGAACATGCTGGAGATCTTGGTTTCAA 856
265 sValProGlnAlaGlnProAlaLysProAlaLeuLysSerGluLys 281
857 AGTCCCGCAGCGCA...GCCACCAAAACCCGACTCAAGTCAGAGGAAAG 903
282 ThrProValLysLysProGlyLysGlyArgLysValThrPhePheGlu 298
904 ACACCATTAAGAAACAGGAGGATGAGAAAGTAACCTTTTGTGAAC 953
298 roSerProGlyLysPheGluAsnGlyThrSerAsnLysAspGluPheAr 314
954 CTGGCTCTGGGATGGAATAATGGACTAGTAATAAAGAGATGAGTTTCA 1003

```

```

314 gmetProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuProme 331
1004 GATGCCCTTATCTAAGTATCAGACGCTGCTGGAATTTCTCCATGG 1053
331 aLProGluValAlaGlnAlaValAlaGlyValSerGlnLysHisHisThr 347
1054 TGCCGAGAGTCCGCCAGGCTGTAAGAGTTAGTCAAGACATCAGCAAA 1103
348 AspPheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAla 364
1104 GATTTTACAGGGGACACTCCGAAATCTCGCAAGCCACGGAATCTCCAT 1153
364 tIleAlaArgGluLeuLeuTyrGlyThrSerProThrAlaGluThr 381
1154 GATAGCCCGAGAGTTGTTATGGGGCACCTGCCCCACAGCCGAAACCA 1203
381 LeuLysSerAsnLysSerSerGlyHisValProHisGlyProArgThr 397
1204 TTTTAAAGATATACATCTCTCAGGCCAGGTACCCATGAGACTCTCAG 1253
398 ArgProSerGluGlnLeuTyrTyrLeuSerArgAlaGlnGlyPheLys 414
1254 AGACCTCTGAGCACTGAGACTATTTCCAGAGTCCAGGATTCAGGT 1303
414 lGluTyrLysAspPheProLysAsnAsnLysAsnGluCysValSerLeu 431
1304 TGAAATCAAAAGACTTCCCAAAAACAAAGAAAGAAATTTGTTATCTT 1353
431 LeuAsnCysSerSerGlnProProLeuValSerHisGlyIleGlyLys 447
1354 TCATATGCTCTCTCTCAGCCACCTGTCATCAGCATGATTCGCAAGAT 1403
448 ValGluSerCysHisAspMetAlaAlaLeuAsnIleLeuLysLeuLeu 464
1404 GTGGAGTCTCTCCATGATATGCTGCGCTGAATCTTAAAGTTGCTGTC 1453
464 rGluLeuAspGlnGlnSerThrGluMetProArgThrGlyAsnGlyPro 481
1454 TGAGTTGGACCAACAAATGACAGATGCCAAGAACAGGAAACGACCAA 1503
481 aISerAlaCysGlyArgCys 487
1504 TGTCTGTGTGTGGAGGTGC 1523

seq_name: gb_htgl4:AC063967

seq_documentation_block:
LOCUS AC063967 225016 bp DNA HTG 06-SEP-2000
DEFINITION Mus musculus clone RP23-41G23, WORKING DRAFT SEQUENCE, 13 unordered
pieces.
ACCESSION AC063967
VERSION AC063967.3 GI:9972323
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus.

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 225016)
Mccombie, W.R., Baker, J. P., Bahret, A., Bal, H., Dedhia, N.N., de la
Bastide, M., Huang, E.N., King, L., Kitchoff, R.A., Miller, B.,
Nascimento, L.U., O'Shaughnessy, A.L., Preston, R.R., Rodriguez, M.A.,
Shah, R.S., Sheker, M., Spiegel, L.A., Toch, K. and Vil, M.D.
Mouse genomic sequence
Unpublished
2 (bases 1 to 225016)
Mccombie, W.R.
Direct Submission
Submitted (22-APR-2000) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
On Sep 6, 2000 this sequence version replaced gi:8439872.
COMMENT
----- Genome Center

```



88193 GCACATGTGAAGTGCATGCGACGAAACACTGCATCAGAAAGATAGGAA 88144  
186 ..... 186  
88143 TCTTTATTCAGACATTGGCCCTTTTCCGACATGAAAGGAGCAGACATT 88094  
186 ..... 186  
88093 CAGCCAACTCTCTCTTTTTCATCTGTGTGTCTCTGAATAATTTACTTCATC 88044  
186 ..... 186  
88043 ATCCCTGACAGTGCATGCGGGAAGAGAGGCCCTAGTCTCTCAGAA 87994  
186 ..... 186  
87993 GCTTGAGTCTCTCTCCCGGCTCACCTGAGGGTTTAAATTCATA 87944  
186 ..... 186  
87943 ACAGTGTGTTGTTTGTCTACTGCGGATGGATGAGAGACTCAGTGTGC 87894  
186 ..... 186  
87893 TGACCAAGCAAGGCTCTTCATAGGATGATCTTCACACTGAGGCTT 87844  
186 ..... 186  
87843 TACAAAGCTTTAATGTGTGTGTTTAAATGAGAAAGCTAGGC 87794  
186 ..... 186  
87793 ATGGGTCCCATGTATGATCAGTACGAGACAGTCAAGCCGTGTTCAAT 87744  
186 ..... 186  
87743 CTTCCCTTTTGTCAATCTTTTTTTTTTTTTTCTGTGAACTCACTCT 87694  
186 ..... 186  
87693 GTAGACGAGGCTGGCTCGAATTCAGAAATCGGCTGCCCTCCACCTCCA 87644  
186 ..... 186  
87643 AGTCTGGGATTAAAGCGGTGCACACACACTGCCGCTCTTTTGTCA 87594  
186 ..... 186  
87593 TTCTGCCAGTGTGCTTACAGCTTTTTCATGCGAGTTCCACTGTGC 87544  
186 ..... 186  
87543 TAATGCCCACTTACATGAGTCACTTCGCAATCCGAGAGTACAGAA 87494  
186 ..... 186  
87493 GGTACAGGCAACCAACCAAGAGACATCTCAGCATGATCTCTGAAC 87444  
186 ..... 186  
87443 ATGTTTAGCTTAGAGAGGTTTGTCTGAGCGTTTACACAGACATA 87394  
186 ..... 186  
87393 GGATGTTTCTGTCTCAGTCCGTGCTGACGCTGACTGCTGTCTC 87344  
186 ..... 186  
87343 TCTGTGAGTAAACAGCAATGCTCTGCACGTGCTTTTTCAGCTACAG 87294  
187 TTTAATAPrAsPTyrgLyInglymeAsnProIleSerArgLeuIaGl 203  
|||||

87293 ACAGCCCGGATTATGCGCAAGGATGAATCTATTTAGTACTTCACA 87244  
203 nllieInglInAlaLySgluLySgluProgluTyMetIleuThg 220  
|||||  
87243 GATCCAGCAGGCAAAAAGAGAGAGCCAGATCATGCTCTTACAG 87194  
220 luArgLyLeuProArGArGArgGluPheValMet..... 231  
|||||  
87193 AACGAGTCTTCCAGCTCCGAGGACTTGTGATGCAAGTGAAGTGTGTG 87144  
231 ..... 231  
87143 GGTGTAGCAGCCCTTCCCTGACATGCTTAATCCGCGAGTGTGACAG 87094  
231 ..... 231  
87093 TTAGCTGACTTGTAGTGCCTGCTCTGTGCTCTCTGAGGCCAAGGCAA 87044  
231 ..... 231  
87043 GCGCAGGTGATCTCTATGATTCAGACTCCACATTTAGTGATCCCA 86994  
231 ..... 231  
86993 TCACAGAAACAGGATCCACTCTGTAAAGTGAAGTGGACCAGATTGCG 86944  
231 ..... 231  
86943 TGAAGTGGCTGGGGGACACCTTTGATGAAGTGTCTGCAATTTGTCCAGC 86894  
231 ..... 231  
86893 AGCCAGCCCTCAGAGAGTGTATGAGAGGAGAGCCCACTGCACCTAG 86844  
231 ..... 231  
86843 AATATATCCTGTCTTAATAATTTGATGATGTGTGTGCACACATTTTC 86794  
231 ..... 231  
86793 CATACATAGTTTAAAGTTTAAAGAACTCTTAAAGCAGTGTAGAT 86744  
231 ..... 231  
86743 GGCTCAGTGGTTAAGACACTGACTCTCTTCGAAAGTCTGTAGTTCAA 86694  
231 ..... 231  
86693 ATCCAGCAGCACATGTGTCTCACACACCACCAATATGAGATCTGATG 86644  
231 ..... 231  
86643 CCCTCTTGTGTGCATCTGAAGACAGTACAGTACTTAATAAGTCT 86594  
231 ..... 231  
86593 TTTAAAAAATAAATCTTTTAAATATAGAGTTTCAATATCTCTAGCT 86544  
231 ..... 231  
86543 GTGTGTGCCATCACAGAGTAGTACTCAGAAATAGAAATAAAGCTT 86494  
231 ..... 231  
86493 TTTACCATCAAGTTCAGAGTATCCCTGGGCTTTGGTTAAGTACAG 86444  
232 ..... 232  
86443 GTCTTTGTTGTGTTTACACAGGTAAAGTTGGCATACACTGCAGAA 86394  
242 GlyValGlyThrAsnLySValAlaLySArgAsnAlaGluAsnMe 258  
|||||  
86393 GGAGTGGTACCAATTAAGAGGTGGCCAAAGCGTAATGCTGTGAGAACAT 86344

```

258 LLeuGIuIleLeuGIyPheLysValProGIuAlaGIuProAlaLysProA 275
|||||
86343 GCTGGAGATCTGGGGTTCAAAAGTTCCCAAGGCGAGCTGCCAAAGCAG 86294
275 lLeuLysSerGIuGIuLys..... 281
|||||
86293 CACTCAATCAGAGAGAGAGGTGACTGTGGACTGTGTGCCAGGTGGCC 86244
281 ..... 281
86243 TAACTCCCTGCATGGCAGAGCTGGGGCAAGGCTTCACTCCCTGCAGT 86194
281 ..... 281
86193 AAGCGGCTGCCCCCGGGGCTCCTCCATGAGATTAGATTGGCAGGAGG 86144
281 ..... 281
86143 GAGTTTCGTGCCAGTGTGCACATCTGGAGGCAATTCATTCGTTAAG 86094
281 ..... 281
86093 TCTCCTGTTAGTGTCACTATGGGCTAGATGTGCATCTCCGAGAGGTC 86044
281 ..... 281
86043 CTGCTCTACTTCTGATGGTACTTGGGGTTGGAGGGAACATAAATGAC 85994
281 ..... 281
85993 TGGCACCCTGTGCATGACCCAGTTTGAATGCTCTGTGGTTTCCGCG 85944
281 ..... 281
85943 GACAGCTAGCAGTGTCAATAGCCAAATCTATGATTGAGTTGAATA 85894
281 ..... 281
85893 TCACATACAGCACCATTGCTGGCTACAGCAGTTGTCAATTTGGGGATTG 85844
281 ..... 281
85843 AGCTTCTGAGCTGTGCCAGTTTCATTACAAGAAATAGTCCCTCCACT 85794
281 ..... 281
85793 GCTGGAGCTTAGTACTGTGCTGCTGATTTGGCTTCCAGATTTC 85744
281 ..... 281
85743 CATTCCAAACAGTGTAGAGTCTATTAGTTCAACACTCCACATAGTGAGA 85694
281 ..... 281
85693 TCCCGTACAGAAACAGTATCCGCTCTGTAAGATGAGTTGGACAGAGT 85644
281 ..... 281
85643 TTGGCTGAGTGTGGGGGACAGCTTTGATGAAGTGTCTGAATCTGT 85594
281 ..... 281
85593 CTCTTCTGTGTATGACATTAGGTACCTGTGTCTTCTTGTGATGTAC 85544
281 ..... 281
85543 TGTTTTGGATGTTCTAGTTATGTAAGTGTCTGCTAGCACCAGTGC 85494
281 ..... 281
85493 AGACCTACTTTCCCAACAAAGCTGACACTTTTAAGCAGCATTTTCTCC 85444

```

```

281 ..... 281
85443 TGGCAGTGTGTTCTGTGTTCTGGACAGTGCACAGTATGATTCAGAT 85394
281 ..... 281
85393 CTGCTGTGTCAAAATGCTGCCATGGGGAGGCGAGCGCTTGTGTATCCA 85344
282 ..... 282
85343 CACAGACCCCGAGTGTCTATGCTCAGCTCCATTTTCTTCTTAAAGA 85294
282 hrProValLysLysProGIyAspGIyArgLysValThrPhePheGIuPro 298
|||||
85293 CTCAGTAAAGAAACACAGAGACGGAAGAAAGTAAGCTTTTGGAACT 85244
299 SerProGIyAspGIuAsnGIyThr..... 306
|||||
85243 AGCCCTGGGATGAATAAGTAAGTGTGGCCCTCTGTCAATTC 85194
306 ..... 306
85193 AACCAATGCTCTGTGCTGTGTGCTCTGTGAGTTTGGAAACACA 85144
306 ..... 306
85143 GTATTCTGCTCTCTGTGTATGCTGGGATGTAAGGTGCTGTAGC 85094
307 ..... Ser..Asn 308
85093 TTCTTGGGTGTGATGTGACTTGTGTGACCTGTGTCCCTGCAGGTAC 85044
309 LysAspGIuGIuPheArgMetProTyrLeuSerHisGIuGIuProAl 325
|||||
85043 AAGGCGAGAGTTTCAGATGCTTATCTTACCCCTGCAGCTGCCAGC 84994
325 agIyIleuPromeTValProGIuValAlaGIuAlaValGIyLysSerg 342
|||||
84993 TGAATTTCTCCCATGTGTGCGGAAAGTTGCCAGGCTTCGGGGTTATTC 84944
342 IngIyHisHisThrLysAspPheThrArgAlaAlaProAsnProAlaLys 358
|||||
84943 AAGGACACACACCAAGATTTCACACAGGCGAGCTCCAAATCTGCCAAG 84894
359 AlaThrValThrAlaMetIleAlaArgIleuLeuTyrGIyLysThrSe 375
|||||
84893 GCAAGGTAACTGCCATGATAGCCCGAGAGTTGTGTAGCGGGGCACCTC 84844
375 rProThrAlaGIuThrIleLeuLysSerAsnIleSerSerGIyHisValP 392
|||||
84843 GCCCAGCCGAGAACCATTTTAAGAGTAACATCTTCAAGCCACGTAAC 84794
392 rHisGIyProArgThrArgProSerGIuGIuLeuTyrTyrLeuSerArg 408
|||||
84793 CCCATGAGCTTCGCAGTGTAGACCTCTGTAGCAACTGTACTACTTCCAGA 84744
409 AlaIngIyPhe.Gln..... 413
|||||
84743 GCCCAGGATTCAGGTACCGTGTGTCAATCTGTGGCCTTACGTTCTT 84694
413 ..... 413
84693 CCTGCACAGAAATGCTTGTGTAGTGTGCAAGAACCTGGGTTCTGTCT 84644
413 ..... 413
84643 CAGCACTACAGAAACAGAAAAAGCAAGATTATATGCTTTATTAAGATT 84594
413 ..... 413
84593 TGAATGTACAGTAAGTTCTGAGAGGAAGAAATGTCTGGGATGAGAAATG 84544
413 ..... 413

```

```

84543 GCTTATATAAGGTAGTATTCTTAAGCCCAACAACCTTGATATGATTA 84494
413 ..... 413
84493 CTGGTGGAAGAACCAACTCCAGCAAGTTGTCTCACTGTGCACACCCTG 84444
413 ..... 413
84443 CCCAGAGAGTTTACGTGTGGCCAGGCTGAAGGAGAGAACACCTTTT 84394
413 ..... 413
84393 ACTTAGGAACACACACTACATGAGCATGGAGAAGAGCTTGAGGGTTC 84344
414 .....ValGluTyr 416
84343 TGGGCTGGGGCCACATGCAAGAGTGTCTTTTCTTCCATTAGTTGATA 84294
416 TLYSASPheProLysAsnAsnLysAsnGluCysValSerLeuIleAsn 433
84293 CAAGATTTTCCCAAGACAACAAGAGAGTGTATCTCTCATCACT 84244
433 ySerSerGlnProProLeuValSerHisGlyIleGlyLysAspValGlu 449
84243 GCTCTCACAGCCGCTCTGCTCAGTCATGCGCAAGATGTGGAG 84194
450 SerCysHisAspMet..... 454
84193 TCCTGTCATGATATGTATGTGGGCCAGAGAGTGTGAGAGAGCTGGC 84144
454 ..... 454
84143 ACATTCCTCTTGTTCAGTCAGTGTGTAGTGCCCTCAGAAAGAGGACATG 84094
454 ..... 454
84093 GTTCTCCTGAGCCTTGTGACATGTCTGATTCATTTCATTCTTCT 84044
454 ..... 454
84043 TCCTGCTTCTTCTTCTTCTTCTCCTCCTGAGAGCTGTAGAAAGCCCC 83994
454 ..... 454
83993 TTTGGGAGTTCGTTCACAAGAAGATCAGACACTTGAGCTTTGGTGT 83944
454 ..... 454
83943 TCACCTGATGAGACATATGACACCTTTAAGCGGGACCCCTGATTTG 83894
454 ..... 454
83893 AAGGAACATGGAGTGTGATGTGAGCAGATTCAMACATTCATTG 83844
454 ..... 454
83843 GAAGTTTAAAGCTGAGCTAGCTGTCTGCTGCTGTTGATTATTCCT 83794
455 .....Ala..AlaLeuAsnIleLeuLysIleLeuSerGlnL 466
83793 TCCTTACCTTTCTAGGCTGCACTGAACATTTTAAAGCTGCTGTGAGT 83744
466 euAspIleGlnSerThrGluMetProArgThrLysAsnGlyProValSer 482
83743 TGGACCAACAGACAGAGATGCAAGAACAGAAATGAGCAAGTTTCA 83694
483 Ala 483
83693 GCC 83691
seq_name: gb_pr6: HSA9062

```

```

seq_documentation_block:
LOCUS      HSA9062      4058 bp      mRNA      PRI      14-JAN-2000
DEFINITION Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominiidae; Homo.
ACCESSION  Y19062
VERSION    Y19062.1 GI:6634456
KEYWORDS   39k3 protein.
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 4058)
            Buchner,G., Bassi,M.T., Andolfi,G., Ballabio,A. and Franco,B.
            Identification of a novel homolog of the drosophila staufen protein
            in the chromosome 6p (1), 113-118 (1999)
JOURNAL    Genomics 62 (1), 113-118 (1999)
MEDLINE    20054364
REFERENCE  2 (bases 1 to 4058)
            Franco,B.
            Direct Submission
            Submitted (07-JUN-1999) B. Franco, TIGEM, Via Olgettina 58, 20132
            Milan, ITALY
FEATURES   Location/Qualifiers
            source          1..4058
                                /organism="Homo sapiens"
                                /db_xref="taxon:9606"
                                /clone_lib="teratocarcinoma-neuron"
                                /map="8q13-8q21.1"
                                202..1642
                                /codon_start=1
                                /product="39k3 protein"
                                /protein_id="CA664341.1"
                                /db_xref="GI:6634457"
                                /translation="MLQINQMEFVSLGSDQWSESSSIKKAQAVANKALTESTLP
                                KPVQKPKSNVNNNGSITPTVELNLGMLKNGEPIATPLDPPPPRANRANRPNQY
                                NQRYHCPVKLIPYQVLTVGNNEFEGKTRQARNAAMKALQALQNEPIPERSPQNG
                                ESGKMDMDKDKANSSEISLVEIALKRMNPVSFEVLEKSGPPHMSFVTRVSGEFSF
                                EGGNSKLSKRRATVTLQELKLPPLVVEKPLPVEKPKRTIVKAGPEYGGQMP
                                ISRLAIOQAKKEKPDVLLSERGMPPRRFVQVKGYNVATGCGNKKIAKNAA
                                EAMLLQIGKASTNLQDLEKTEGKNGKSGSPKPGPPPTNTPGIIILSPDVQEME
                                ASRKIVISGTTLGLISPKDMNPSSPSISPTSSATIRALLNQTSTAEALIGL
                                KGSPTPCSPVQPSKQLEYLARIGFQV"
BASE COUNT 1280 a 731 c 737 g 1310 t
ORIGIN
alignment_scores:
Quality: 927.00 Length: 449
Ratio: 2.915 Gaps: 14
Percent Similarity: 70.824 Percent Identity: 48.775
alignment_block:
US-09-316-048-11 x HSA9062 ..
Align seg 1/1 to: HSA9062 from: 1 to: 4058
1 MetTyrLysProValAspProHisSerArgMetGlnSerThrTyrSerTyr 17
   ::::::::::::::::::::::::::::
439 ATCTACAGGCCATTAGATTCACAAGCCATTCCCAATTTATAGACTAATTA 488
   ::::::::::::::::::::::::::::
17 rglMetArgGlyGlyAlaTyrProProArgTyrPheTyrProPhePro 34
   ::::::::::::::::::::
489 CAACCTTTCGG...GGCATGTACATCAGAGGTATCAT.....TGCCACAG 529
   ::::::::::::::::::::
34 alProProLeuLeuTyrGlnValGluLeuSerValGlyGlyGlnGlnPhe 50
   ::::::::::::::::::::
530 TGCCTAGAGATCTTTAT...GTTCAGCTCCTAGCAAAATATATGATTTT 576
   ::::::::::::::::::::
51 AsnGlyLysGlyLysMetArgProProValLysHisAspAlaProAlaAr 67
   ::::::::::::::::::::
577 TTTGGGGAGAGAGAGACTGACACAAGCTGTAGACACATCTCCATGAA 626
   ::::::::::::::::::::
67 galAlaLeuArgThrLeuGlnSerGluProLeuProGluArgLeuGluVala 84
   ::::::::::::::::::::

```







together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX>

RP13-140E4 is from the library RP13-13.1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/> VECTOR: pBAC3.6

This sequence is the entire insert of clone RP13-140E4.

# FEATURES

## source

```

1. .163253
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="X"
   /clone="RP13-140E4"
   /clone_lib="RP13-13.1"
   25. .188
repeat_region
   /note="L1M4 repeat: matches 5114. .5296 of consensus"
184. .411
repeat_region
   /note="L1P3 repeat: matches 2841. .3072 of consensus"
412. .694
repeat_region
   /note="AluSq repeat: matches 1. .287 of consensus"
695. .3450
repeat_region
   /note="L1P3 repeat: matches 232. .2841 of consensus"
3451. .6427
repeat_region
   /note="L1Pa2 repeat: matches 3148. .6144 of consensus"
6432. .7143
repeat_region
   /note="L1MD1 repeat: matches 5283. .6016 of consensus"
7164. .7316
repeat_region
   /note="Aluub repeat: matches 137. .290 of consensus"
7339. .7646
repeat_region
   /note="L1MD2 repeat: matches 6016. .6330 of consensus"
7650. .7754
repeat_region
   /note="L1MC5 repeat: matches 7608. .7714 of consensus"
7767. .8072
repeat_region
   /note="AluY repeat: matches 1. .305 of consensus"
8095. .8136
repeat_region
   /note="21 copies 2 mer ca 90% conserved"
8152. .8245
repeat_region
   /note="L1MC5 repeat: matches 7478. .7572 of consensus"
8273. .8494
repeat_region
   /note="Alufo repeat: matches 84. .305 of consensus"
8503. .8633
repeat_region
   /note="L1MC5 repeat: matches 7305. .7446 of consensus"
8712. .8879
repeat_region
   /note="84 copies 2 mer tt 72% conserved"
8899. .9190
repeat_region
   /note="Alufo repeat: matches 8. .300 of consensus"
complement(9869. .10121)
   /note="match: STS: Em:L41960; match: STS: Em:L41960"
12531. .13017
repeat_region
   /note="L2 repeat: matches 2083. .2505 of consensus"
13018. .13320
repeat_region
   /note="L1M1 repeat: matches 6004. .6303 of consensus"
13321. .14004
repeat_region
   /note="L2 repeat: matches 1197. .2083 of consensus"
complement(16833. .17221)
   /note="match: GSS: Em:A058019"
misc_feature
```

```

16840. .17277
   /note="match: GSS: Em:B65742"
   complement(16845. .17241)
   /note="match: GSS: Em:A0149727"
   complement(16848. .17220)
   /note="match: GSS: Em:A0240745 Em:A0474471"
16886. .17629
repeat_region
   /note="MER49 repeat: matches 234. .973 of consensus"
18183. .18407
repeat_region
   /note="L1P3 repeat: matches 5928. .6148 of consensus"
18455. .19002
repeat_region
   /note="MER41A repeat: matches 1. .554 of consensus"
19289. .19465
repeat_region
   /note="MIR repeat: matches 15. .198 of consensus"
19583. .20000
   /note="match: GSS: Em:A0017861"
19875. .20274
repeat_region
   /note="MER57B repeat: matches 1. .403 of consensus"
20275. .21084
repeat_region
   /note="MER57-internal repeat: matches 1. .806 of consensus"
21078. .22422
repeat_region
   /note="MER57-internal repeat: matches 440. .2799 of consensus"
22425. .22764
repeat_region
   /note="MER57-internal repeat: matches 3892. .4230 of consensus"
22765. .23041
repeat_region
   /note="AluSq repeat: matches 8. .286 of consensus"
23042. .24919
repeat_region
   /note="MER57-internal repeat: matches 4230. .6114 of consensus"
24939. .25118
repeat_region
   /note="L1P1 repeat: matches 5980. .6153 of consensus"
25119. .25422
repeat_region
   /note="AluY repeat: matches 1. .304 of consensus"
25423. .25607
repeat_region
   /note="L1P1 repeat: matches 5777. .5980 of consensus"
25635. .25672
repeat_region
   /note="19 copies 2 mer ta 78% conserved"
25681. .26070
repeat_region
   /note="L1P1 repeat: matches 5397. .5790 of consensus"
26082. .27136
repeat_region
   /note="MER57-internal repeat: matches 6489. .7537 of consensus"
27137. .27249
repeat_region
   /note="MER57B repeat: matches 1. .114 of consensus"
27250. .28097
repeat_region
   /note="L1MC3 repeat: matches 6763. .7631 of consensus"
28123. .28390
repeat_region
   /note="MLT1C repeat: matches 28. .313 of consensus"
28404. .28952
repeat_region
   /note="L1MC3 repeat: matches 5376. .5993 of consensus"
28946. .29622
repeat_region
   /note="L1M4 repeat: matches 4880. .5566 of consensus"
30851. .31163
repeat_region
   /note="AluSc repeat: matches 1. .301 of consensus"
32518. .34668
repeat_region
   /note="L1PA14 repeat: matches 3983. .6149 of consensus"
34672. .35189
repeat_region
   /note="MIR2B repeat: matches 1. .440 of consensus"
35190. .36499
repeat_region
   /note="L1PA14 repeat: matches 2718. .3988 of consensus"
36500. .36876
repeat_region
   /note="THE1C repeat: matches 1. .369 of consensus"
36877. .38213
repeat_region
   /note="L1PA14 repeat: matches 1264. .2718 of consensus"
38221. .39169
repeat_region
   /note="L1PA14 repeat: matches 288. .1240 of consensus"
39455. .39517
repeat_region
   /note="L2 repeat: matches 2349. .2404 of consensus"
40251. .40337
repeat_region
   /note="MIR repeat: matches 59. .146 of consensus"
41354. .41648
repeat_region
```

```
/note="AluX repeat: matches 1..295 of consensus"
1650..41743
/note="47 copies 2 mer aa 62% conserved"
repeat_region 41744..42318
/note="L1 repeat: matches 4553..5128 of consensus"
repeat_region 42317..42391
/note="Alu repeat: matches 251..302 of consensus"
repeat_region 42360..42395
/note="Alu repeat: matches 79..114 of consensus"
repeat_region 42421..43734
/note="L1 repeat: matches 4019..5334 of consensus"
repeat_region 43735..43834
/note="FLAM3C repeat: matches 32..131 of consensus"
repeat_region 43835..44621
/note="L1 repeat: matches 3232..4019 of consensus"
repeat_region 44622..44802
/note="L1 repeat: matches 6121..6304 of consensus"
repeat_region 44803..45815
/note="L1 repeat: matches 2126..3232 of consensus"
repeat_region 45514..46762
/note="L1 repeat: matches 1155..2794 of consensus"
repeat_region 46760..47327
/note="L1 repeat: matches 59..642 of consensus"
repeat_region 47325..48451
/note="L1 repeat: matches 1334..154 of consensus"
repeat_region 48533..48727
/note="L1 repeat: matches 4030..4226 of consensus"
repeat_region 48717..49410
/note="L1 repeat: matches 5092..5805 of consensus"
repeat_region 49411..49506
/note="48 copies 2 mer at 69% conserved"
repeat_region 49528..50023
/note="L1 repeat: matches 5787..6308 of consensus"
repeat_region 50050..50333
/note="Alu repeat: matches 1..300 of consensus"
```

```
alignment_scores:
Quality: 705.50 Length: 446
Ratio: 2.383 Gaps: 15
Percent Similarity: 66.366 Percent Identity: 43.722
```

## alignment\_block:

US-09-316-048-11 x HSB140E4

Align seg 1/1 to: HSB140E4 from: 1 to: 163253

```
1 MetTyrLysProValAspProHisSerArgMetGlnSerThrTyrSerTy 17
.....
159120 ATCCACAGGCCCTTAGATCCCAAGCCATTCCCAATATAGAGCTAATT 159169
17 rGlyMetArgGlyAlaTyrProArgTyrPheTyrProPheProv 34
.....
159170 CAACCTTCAA...GTTCATGTATAATCAGAGTATCATTTGCCAATTA 159212
34 aLProProLeuLeuTyrGlnValGluLeuSerValGlyGlnGlnPhe 50
.....
159213 .....CTTAAGATCTTTATGTTCAGCTCAT..GTAGAAATATAGATTT 159256
51 AsnGlyLysGlyLysMetArgProProValLysHisAspAlaProAla 67
.....
159257 TTGGGGGGAAGAAAGCC.....AGACACAATGCTGCATATGA 159293
67 gAlaLeuArgThrLeuGlnSerGluProLeuProGluArgLeuGluVal 84
.....
159294 AGACATCCACAGACATGACATTTGAACCTATTCCAGAAAGATGCGCTTAG 159343
84 sngLYArgGluAla.....GluGluGluAsnLeuAsnLysSer 96
.....
159344 ATGGATGATCAGCAAGAGATATAGATGATGACAAAGATCAATATAGCCT 159393
97 GluIleSerGlnValPheGluIleAlaLeuLysArgAsnLeuProValAs 113
.....
159394 GAGATCAGCTTAGTGTTCGAATATGCTGTAAGTGAATATATACCTGTTAG 159443
```

```
113 nPheGluValAlaArgGluSerGlyProProHisMetLysAsnPheVal 130
.....
159444 TTCTGAGGTTATTATTAAGAAAGTGACACCAATATGAAAGCTTTGTTA 159493
130 hArgValSerValGlyGluPheValGlyGluGlyLysSerLys 146
.....
159494 CTCGAGTGTACAGTAAGAAAGTTCTGTACAGAAAGCAAGAAATAGCAAA 159543
147 LysIleSerLysLysAsnAlaAlaArgAlaValLeuGluGlnLeuArg 163
.....
159544 AAACCTCCCAAGAGCGTGTCCAAACACCGCTTACAGAGCTTAAAAA 159593
163 gLeuProProLeuProAlaValGluArgValLysProArgIleLysLys 180
.....
159594 ACTTCCACCTCTTCTGCTGTGCGAAAGCAAAATATTATTTAAAAAAC 159643
180 ySerGlnProThrCysLysThrAlaProAspTyrGlyGlnLysMetAsn 196
.....
159644 GCCCTAAACAATAGTACAGCGCGACAGAAATA..GGCCAAGGATGAGC 159692
197 ProIleSerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGlu 213
.....
159693 CCTATTAGCTGCTGCGCCCAATTCACAGGCCCAAAAGAAAGAGAGCC 159742
213 oGluTyrMetLeuLeuThrGluArgGlyLeuProArgTArgGluPhe 230
.....
159743 ACATTAATGTTTGTCTTCAGAAAGAGAAATGCTCATGTCCAGATTTG 159792
230 aMetGlnValLysValGlyHisHisThrAlaGluGlyValGlyThrAsn 246
.....
159793 TGATGTATGTGAAGGTAGCAATGAAGCTGTACAGAACGGACCTAAT 159842
247 LysLysVal...Ala..LysArgAsnAlaAlaGlnAsnMetLeuGluIle 262
.....
159843 AAAAAGGCCAAAAAAGTGTTCAGTACAGCAATGCTATTATTAACAC 159892
262 euGlyPheLysValProGlnAlaGlnProAlaLysProAlaLysSer 278
.....
159893 TTGGTTATTA.....GCATCCATTAAATCTTCAGAT 159924
279 Glu...GluLysThrProValLysLysProGlyAspGlyArgLysVal 294
.....
159925 CAACCTGAGAGACA.....GGGGAACAACAAG 159953
294 rPhePheGluProSerProGlyAspGlnAsnGlyThrSerAsnLysAsp 311
.....
159954 ATGAGGTGATCCACAGCGCTGGGTTCTGTGACCAACCAAAATAT 159996
311 LuGluPheArgMetProTyrLeuSerHisGlnGlnLeuProAlaGly 327
.....
159997 .....ACTCCAAAGCAATT 160011
328 LeuProMetValProGluValAlaGlnAlaValGlyValSerGlnGly 344
.....
160012 CTTCATTTGTCTCCGATGATGTTATCAAGAGATGAAAGCCAGCACAA 160061
344 sHisThrLysAspPheThrArgAlaAlaProAsnPro..... 356
.....
160062 AGTAACCTCTGACACTACTAGGCTATTGTCACCCAGAGATATGAAAC 160111
357 .....AlaLysAlaThrValThrAla 363
.....
160112 AACCTCAAGCTCTTTGTCAGTATGATCCACATCAATATGTTACAGT 160161
364 MetIleAlaArgGluLeuLeuTyrGlyGlyThrSerProThrAlaGlu 379
.....
160162 ACAATGGCCAGGAACTCCTTATGAATGAAACATCTTTACAGCTGAAC 160211
380 ThrIleLeuLysSerAsnIleSerSerGlyHisValProHisGlyPro 396
.....
160212 CACAGGTTTAAAGGA.....AGTTCTCTACTGCCCTTGTTCGA.. 160254
```

```

396 rgtTharProSerGIuInLeuTYrTleuSerArGAlaGInGlypHe 412
      ::::::::::::::::::::|
160255 ..GTACAACCTTCAAAACACTGGAATATTACGAAGATTAAGGCTTT 160302
      413 GInVaIGluTYrTlysAspHeProLysAsnAsnLys 424
      ||||| |:::| |:::| |:::| |:::|
160303 CAGGTATGAATTAAAGC.....AAAACCAAAAA 160332

seq_name: qb_pr4:AL133174

seq_documentation block:
LOCUS      AL133174      113836 bp      DNA      PRI      14-JUN-2000
DEFINITION Human DNA sequence from clone RP3-470L14 on chromosome 20, complete
sequence.
ACCESSION  AL133174
VERSION    AL133174.15  GI:8573761
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 113836)
AUTHORS   Ramsay,H.
TITLE     Direct Submission
JOURNAL   Submitted (14-JUN-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB30 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
            requests: clonerequests@sanger.ac.uk
            On Jun 20, 2000 this sequence version replaced gi:8388434.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em.: EMBL; Sw.: SWISSPROT; Tr.: TrEMBL; Wp.: WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone configs of human
            chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr20
            RP3-470L14 is from the library RPCI-3 constructed at the Roswell
            Park Cancer Institute by the group of Pieter de Jong. For further
            details see http://dacpac.med.buffalo.edu/
            VECTOR: pCYPAC2
            IMPORTANT: This sequence is not the entire insert of clone
            RP3-470L14. It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true right end of clone RP3-470L14 is at 113836 in this
            sequence. The true right end of clone RP1-155G6 is at 100 in this
            sequence.

FEATURES             Location/Qualifiers
     source           1..113836
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /clone="RP3-470L14"
                     /clone_1lb="RPCI-3"
                     /note="Weak data"
     misc_feature     95583..95627
                     /note="Single clone region. pUC subcloned to give 12x
                     coverage. Assembly confirmed by restriction digest data."
     BASE COUNT      33974 a 25115 c 24233 g 30514 t
     ORIGIN

```

```

alignment_scores:
    Quality: 701.00      Length: 429
    Ratio: 4.124        Gaps: 3
    Percent Similarity: 39.627      Percent Identity: 37.762

alignment_block:
US-09-316-048-11 x AL133174/rev ..

Align seg 1/1 to reverse of: AL133174 from: 1 to: 113836

281 LysThrProValLysLysProGlyAspGlyArgLysValThrPhePheG1 297
||||| ::::::::::::::|
38483 AAGCAACCCATTAAGAAACCAAGGAGATGAAGAAAGTAACCTTTTGA 38434
297 uProSerProGlyAspGluAsnGlyThr..... 306
||||| ::::::::::::::|
38433 ACCTGGCTCTGGGGATCAAAATGGAGCTAGTAAGTGATCTTAATATCG 38384
306 ..... 306
38383 TCCTCTCTGACGTGATCTGGGATTGCTGATGCTCAATATACAGCA 38334
306 ..... 306
38333 GAGACCCAGCCAGTGAAGTCCCGATGCAAAAGTCTTTATCATTCCTCG 38284
306 ..... 306
38283 ATTCCCTGCGCTGCCAATGCGTGCCAAAGGGAGGCGCGTGGGGTTTGA 38234
306 ..... 306
38233 TGGTGATCATTCGCTTTTCCAGCGCTGTGATGTGATTCGTGTTACCGT 38184
307 ..... 307
38183 GTATGTCCCTCCCGAGTAATTAAGAGATGATGATTCAGGATGCCCTATCTA 38134
||||| ::::::::::::::|
319 SerHisGlnGlnLeuProAlaGlyIleLeuProMetValProGluValAl 335
||||| ::::::::::::::|
38133 AGCATCAGCAGCAGCTGCTGGAATCTTCATGCTGCGCCGAGTGC 38084
335 agInAlaValAlGlyAlSerGlnGlnHisThrLysAspPheThrArg 352
||||| ::::::::::::::|
38083 CCAGCGCTGTAGGAGTTAGTCAAGCATCACACCAGAAAGATTTTACAGGG 38034
352 laAlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArgGlu 368
38033 CAGCTCCGAATCCCTGCCAAGGCCAGGTAACCTGCAATGATAGCCGAGAG 37984
369 LeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysSer 385
||||| ::::::::::::::|
37983 TTGTTGATGAGGGGACCTCGCCACAGCCAGACCATTTTAAAGATAA 37934
385 nIleSerSerGlyHisValProHisGlyProArgThrArgProSerGluG 402
||||| ::::::::::::::|
37933 CATCTCTTCAGGCCACGATACCCCATGGAAGCTCTCAGAGACCCCTGAGC 37884
402 InLeuTyrTyrLeuSerArGAlaGInGlyPheGln..... 413
||||| ::::::::::::::|
37883 AACTGGACTATCTTTCCAGAGTCCAGGAGATTCCAGTAACGTCTGGCCT 37834
413 ..... 413
37833 GAGCTGTGATGAGCTGTGCTCAGAGCAGATCATTTTGGCTTCTCAGA 37784
413 ..... 413
37783 AATGAACTTTGAACCTTTGACTTTTCTTAATCAAGTTCTGTCCAGG 37734
413 ..... 413

```

```

37733 GATGGGCTGGGAGACGATGAGAGGGTAGAGATAGAGTATTGTA 37684
413 ..... 413
37683 AGCATATGCTGAAGAGAAATATCTTATCTTGTAGTAGATTTCAA 37634
413 ..... 413
37633 CATGCAAGCCACCCCTAAAGCTTTAACCACCTAGACAGTGGTTGAATT 37584
413 ..... 413
37583 TGGATTTCAAAGAACAGACTCATCTGCTGACACTTCTGCTGTAGGA 37534
413 ..... 413
37533 AAGGATATGGCTTCTCCAGCTTTGGGCTTCTGACGACACCTAGGTTT 37484
413 ..... 413
37483 CTGGAAGTCTCTGGGCTGACAGGGCAGGTGTAGTATGGCCACAGTGGG 37434
413 ..... 413
37433 CTGAGTTGAAAAAAGAGCTGTGATCTTGTGGGGCTGAGCTGC 37384
413 ..... 413
37383 AGCCCTCCAGCCACCTCATCTGCTGCTCCCTACCTAGCCTCAGCTTC 37334
414 ..... ValGluTyrLysAspPheProLysAsnAsnLysAsnG 426
37333 TTTTCCCTTAGTGTGAATACAAAGACTTCCCAAAAACAAAGAACG 37284
426 LucysValSerLeuIleAsnGysSerSerGlnProProLeuValSerHis 442
37283 AATTGTATCTCTTATCAATGCTCCTCAGCCACCTCGATGACGCAAT 37234
443 GATLleGlyLysAspValGluSerGysHisAspMet 454
37233 GGTATCGGCAAGATGTGAGTCTCTCCATGATATATG 37198

seq_name: gb_in3:DROSTAUFEIN
seq_documentation_block:
LOCUS DROSTAUFEIN 5237 bp mRNA INV 26-APR-1993
DEFINITION D.melanogaster mRNA, complete cds.
ACCESSION M69111
VERSION M69111.1 GI:158505
KEYWORDS
SOURCE Drosophila melanogaster 0-4 hour embryo cDNA to mRNA.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscocomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 5237)
AUTHORS Johnston, D.S., Beuchle, D. and Nusslein-Volhard, C.
TITLE Stauden, a gene required to localize maternal RNAs in the
Drosophila egg
JOURNAL Cell 66, 51-53 (1991)
MEDLINE 91300552
FEATURES
location/qualifiers
source 1..5237
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/dev_stage="0-4 hour embryo"
<1..5237
275..3355
/codon_start=1
/protein_id="AA13062.1"
/db_xref="GI:158506"
/translation="MQNHVNAARAPRHIRAHNHSHSHAHNHLPRGMEQHLGPELQOO

```

```

OQPPPPQPPRRDLARLHHHLLHAOQOQOQTSNQAANAANAAGAAVHHGNTSNGS
NSNSNSNOMOKTIRQOHLSNNGILNPOPPGPPAPFAPLGNPALAYNLOLPH
PPHMAHLGSLYAAPPHYVMSQAKPAKYNYHNSNANSNGSNNSNYSARAKALITP
YRNOKVYVPVVOEVYTPPEPPYTTNNATNTSTNSYIASEPTOEDTSQKPRETBO
PASADHVSSTNIDTALSNEDTSSSGRGKIDTPMKLVNLRANKITHOYRLEB
RGAADHKTFTVTLMDGEYSADGEFKKKAHLAASNAIEETMKHPPKTRSEEGG
PMRTHIPPTVELNALMKLGORTFYLLDPTQIPPDLSVPEFAGHLLTPGGMQ
PPPPAYALRORLNGFVPIPSOPMHPHFPGCORPPKPSRFALPPLGAVHH
GNGPFPVPPTPKITLTFVGKOKFVGITLOAKHDAARALOVKTQIASSEA
LEDMDDEGDKSPISOVHEIGIKRMTHFVLEBEGAHKKNFTTACTIGSITBEG
CNGKRVSKRAAEKMLVBLQPLPPLTPKOTPLRKIVKTPGKSAARBSVSGTD
GTMOQKPERRRRLNPPKDLIDMDADNPTTKLLOQOTREKERTPELLAKNGNET
ARREFEVMSVSGSTARGTNGSKKLAKRNAOALFELLEVAVYTPNETOSSECT
SATMSAVTAPAVEATAEGKVPVATVPMPGIIILRONKPKPAKRDQIVYKNVES
KEEENAEVAAVEANSNSNSNGSSGSDOATEAASESALNTSTGNTSGVS
NSSVNGANTDGNHNAESKNNTSSNSTNSQSGVYMKPOLLYLSKLDEVNFSDY
PKGNHNEPITLVITLSTHPPQICHEGVKSSSESDAASNAIKLITSLGNNAK"
BASE COUNT 1596 a 1291 c 1204 g 1146 t
ORIGIN Chromosome 55A, right arm.

alignment_scores:
Quality: 634.50 Length: 575
Ratio: 1.995 Gaps: 14
Percent Similarity: 55.304 Percent Identity: 30.435

alignment_block:
US-09-316-048-11 x DROSTAUFEIN
Align seg 1/1 to: DROSTAUFEIN from: 1 to: 5237
3 LysProValAspProHisSerArgMetGlnSerThrTyrSerTyrGlyMet 19
:::||||: ||||| :::: |||
1691 CAGCCCATGATCCGCAT ..... TTCTTCATGAGACC 1722
19 LArgGlyGlyAlaTyrProProArgTyrPheTyrProPheProValPro 36
:::||||: ||| ::|||
1723 TGGTGAGCAGCAGCGTTCTCCGCAAGTTTCATCAGCCGTTGCGTGCAC 1772
36 rGluLeu..... 38
||| |||
1773 CGCCTCGAGGCTCATGTACATCATGACCGCAAGCCCATTTCCATCC 1822
39 .....TyrGlnValGluLeuSerValGlyGlyGlnInp 50
:::||||: ||||| ::|||
1823 GTGCCATCGCACCCAGCAAAATCACCCCTGTTGTGGCAGAGCAAGTT 1872
50 eAsnGlyLysGlyLysMetArgProProValLysHisAspAlaProAla 67
| ||| ||| ::|||
1873 TGTGGGCATCGGACGACACACTGACGACGCAAAACATGACGCTGCGGCA 1922
67 rGluAlaLeuArgThrLeuGlnSerGluProLeu.....ProGluArg 80
||||| ||||| ::|||
1923 GACCGGTGAGGTTCTCAAGACACAGCCCATATCTGCTGAGGAGGCT 1972
81 LeuGluValAsnGlyArgGluGluGluGluAsnLeuAsnLysSerGln 97
||||| ::|||
1973 CTCGAG .....GATTCGATGATGAGGCGGACAAAGAGTCGCC 2010
97 uIleSerGlnValPheGluIleAlaLeuLysArgAsnLeuProValAsn 114
||||| ||||| ::|||
2011 CATCTCGCAGGTTCCAGCAGATCGGATTAAGCCACACATGACCGTTCATT 2060
114 heGluValAlaArgGluSerGlyProProHisLysMetLysAsnPheValThr 130
|::|::| ||||| ::|||
2061 TTAAGGTGTTGGCGGAGGAGGCGCGCATGAAGAACTTTATATAACA 2110
131 ArgValSerValGlyLysPheValGlyGluGlyLysSerLysGly 147
||||| ::|||
2111 GCCTGCATTGTGGATCTATTGTCACCGAAGGAGAAATGGCAAAA 2160
147 sIleSerLysAsnAlaAlaArgAlaValLeuGluGlnLeuArgL 164
|::|::| ||||| ::|||

```

```

2161 AGTGCCAGAAAGCGTCCGCGGAAAAGATGCTGTCAGTTGCCAAAAGC 2210
164 eupProLeuProAlaValGluArgVal..... 173
|||||
2211 TGGCGGCCCTTACGCCACCAAGCAGCGCCATTGAAGCGATTAAAGTT 2260
173 ..... 173
2261 AAGACACCAGCAAAAGTGGTGCAGCTGCCCGAAGAGTTCCGTTGTGTGTC 2310
174 .....LysProArgTLeuLysLysLys 181
|||||
2311 GGGCAGCTGATGACTATGCAAAAGCGGAAAGCGGAGAGACGTAAACGCC 2360
181 erGlnProThrCysLysThrAlaProAspTyrGlyGlnGlyMetAspPro 197
|||||
2361 TCATATCCGCCGAAGACAGCTGATAGACATGAGCGATGCGGATAAATCCC 2410
198 lLaserArgLeuAlaGlnIleGlnAlaLysLysGluLysGluProGln 214
|||||
2411 ATTACAAAGCTAATTCAACTGCACAGCAGCGCTAAGGAAAAGAACCCAT 2460
214 uTyrMetLeuLeuThrGluArgGly.....LeuProArgTArgGlu 229
|||||
2461 TTTCGACCTGATAGCCAAAACCGCATGAACCGCTCGACAGCGGGAAT 2510
229 heValMetGlnValLysValGlyHisHisThrAlaGluGlyValGlyThr 245
|||||
2511 TCGTTATGGAGGTCTCCGCGAGCGGAGCACGCCCTGGAAACAGGCACAC 2560
246 AsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIle 262
|||||
2561 AGCAGAAATTTGGCAAAAGCGAAATGCTGCACAGCGCTATTGTAACGTCT 2610
262 uGlyPheLysValProGlnAlaGlnProAlaLysProAlaLeuLysSerc 279
|||||
2611 G.....GAAGCTGTTCAGTCAACACCAACCAAGCAACAGATCATCGG 2654
279 LuGlu..... 280
|||||
2655 AGGAATGCTGCACTTGGCAACAATGTCGCCGCTTACAGCGCCGCGAGTG 2704
281 .....LysThrProVal..... 284
|||||
2705 GAGGCACTGCGGAAGCAAGTCCCATGTGTACCAACGCCAGTGGGTCC 2754
285 .....LysLysProGlyAspGly 291
|||||
2755 CATGCCAGGCATTTTAATAGTCCGCGCAGAACAGAAACAGCAAAAGAGA 2804
291 rg.....LysValThrPheGluProSerProGlyAspGluAsnGlyThr 306
|||||
2805 GGGATCAAAATTGTAATGTCAGTCAAGTCAATGCAAAAGAGAGAGAG 2854
307 SerAsnLysAspGluGluPheArgMetProTyrLeuSerHisGlnIle 323
|||||
2855 GCCAATAAGAGAGTGGCAGTTGCGAGTGAAGATAGCAAAACAGTGC 2904
323 uProAlaGlyIleLeuProMetValProGluValAlaGlnAlaValGly 340
|||||
2905 AAATCTGTGC.....GACAGCAGCAACAGCAGCAGCGCG 2939
340 aLSerGlnGlyHisHisThrLysAspPheThrAlaGlnAlaProAsnPro 356
|||||
2940 ATTGCAAGCC.....ACAGAGCGGCTAATGAGAGT 2971
357 AlaLysAlaThrValThrAlaMetIleAlaArgGluLeuLeuTyrGly 373
|||||
2972 GCACTGAATACCTCAACCGGAGCAATAGC.....AGTGG 3006
373 yThrSerProThrAlaGluThrIleLeuLysSerAsnLysSerGly 390
|||||
3007 TGTGAGTAGCAACAGCAATGTTGAGCAACACGATGTAAACACC 3056
390 lValProHisGlyProArgThrArgProSer..... 400
|||||
3057 ATGCCGAGAGTAAACACACGAGAGTAGCACCAACAGACAGTAAAC 3106
401 .....GluGlnLeuTyrTyrLeuSerArg 408
|||||
3107 ACACAAAGCGCTGGAGTGCACATGACAGAGCAGCTTGTATCTAGTAA 3156
408 gAlaGlnGlyPheGlnValGluTyrLysAspPheProLysAsnLys 425
|||||
3157 ACTGCTGACTTGTAGGTCAACTCTGCACTATCCGAAAGCAATCACA 3206
425 snGluCysValSerLeuIleAsnSerSerGlnProProLeuValSer 441
|||||
3207 ACGAGTTCCTGACCATGCTGACATTGTCCACACATCCGCGCAGATTGTC 3256
442 HisGlyIleGlyLysAspValGluSerCysHisAspMetAlaAlaLeu 458
|||||
3257 CATGGCGTTGSCAAGAGCTCCGAGAGAGTGCAGAAATGATGCTGCAAGCA 3306
458 nIleLeuLysLeuLeuSerGluLeu 466
|||||
3307 TGCCTTGAAAATCCTCAGCAAGCTG 3331

```









```

|||||
Db 1681 tcaatgcccctctcagccacccctgatcagcatgtaatcgcaagatgtagtccct 1740
Qy 1741 gccatgatatgctgctgcctgcaacatcttaaatgtctgtctgctgagtgacccaagaatga 1800
Db 1741 gccatgatatgctgctgcctgcaacatcttaaatgtctgtctgctgagtgacccaagaatga 1800
Qy 1801 cagaagatgccaagaagaacaggaacagcaaatgctgtctgtgtgagagtgctgaacctttt 1860
Db 1801 cagaagatgccaagaagaacaggaacagcaaatgctgtctgtgtgagagtgctgaacctttt 1860
Qy 1861 ctgagcaatgaacattataaataatcccaacatatatactgaataatctgaactgtctga 1920
Db 1861 ctgagcaatgaacattataaataatcccaacatatatactgaataatctgaactgtctga 1920
Qy 1921 aaatttggaattcttgataccctcccaatgagccgagagacaacagtgagtaagaatgtgg 1980
Db 1921 aaatttggaattcttgataccctcccaatgagccgagagacaacagtgagtaagaatgtgg 1980
Qy 1981 cagcaacagaggaagacaacagaacaacaggaagcgctgtgtgcccgtggaactgtgctg 2040
Db 1981 cagcaacagaggaagacaacagaacaacaggaagcgctgtgtgcccgtggaactgtgctg 2040
Qy 2041 ggggttctgtgtatgagccactcggtgacactggtgcgtccctacgcaatagcaagctgtcgt 2100
Db 2041 ggggttctgtgtatgagccactcggtgacactggtgcgtccctacgcaatagcaagctgtcgt 2100
Qy 2101 gggggaagaagagctgcgccagccagcagctgtctcccgagacaacagatccacacctg 2160
Db 2101 gggggaagaagagctgcgccagccagcagctgtctcccgagacaacagatccacacctg 2160
Qy 2161 ggcacctccgtgtgtgtcttttctccctgtgtgaaagaagaacagccagccacct 2220
Db 2161 ggcacctccgtgtgtgtcttttctccctgtgtgaaagaagaacagccagccacct 2220
Qy 2221 tctcaagctgtgctcactcagacaacattgggaacaacccctgagacagctgcagagagag 2280
Db 2221 tctcaagctgtgctcactcagacaacattgggaacaacccctgagacagctgcagagagag 2280
Qy 2281 ggccttgaccgagccagagcttaaaagcaccagagaagaataatcaatgtcttctctacagc 2340
Db 2281 ggccttgaccgagccagagcttaaaagcaccagagaagaataatcaatgtcttctctacagc 2340
Qy 2341 tgaaccaactttctctagtgtgacagcgccaccacactcctctgagtaaccacacacacac 2400
Db 2341 tgaaccaactttctctagtgtgacagcgccaccacactcctctgagtaaccacacacacac 2400
Qy 2401 caactgtctctctccacaagtgatctgtatcttaattagttcatatttctttgtatga 2460
Db 2401 caactgtctctctccacaagtgatctgtatcttaattagttcatatttctttgtatga 2460
Qy 2461 taagaacatatataaataatcttaattgagaatcttcccaatgtatctagttaaatagcac 2520
Db 2461 taagaacatatataaataatcttaattgagaatcttcccaatgtatctagttaaatagcac 2520
Qy 2521 agttggaactgtctgagacgacttatacaataatccaacgagcaaatgacatctcc 2580
Db 2521 agttggaactgtctgagacgacttatacaataatccaacgagcaaatgacatctcc 2580
Qy 2581 atgtgtatgtgtgtagacatttattatctatctatgaactaacccaggaagcttcagtgatgc 2640
Db 2581 atgtgtatgtgtgtagacatttattatctatctatgaactaacccaggaagcttcagtgatgc 2640
Qy 2641 aaatgtgtgcccctgtgtctgacgtgaacaagctcctggaacttcaaaaacctgtaataag 2700
Db 2641 aaatgtgtgcccctgtgtctgacgtgaacaagctcctggaacttcaaaaacctgtaataag 2700
Qy 2701 tctcccaagctgtataaattggaacatttaggaatttaaaactttagatgacatcttg 2760
Db 2701 tctcccaagctgtataaattggaacatttaggaatttaaaactttagatgacatcttg 2760
Qy 2761 ttccattttatttatttatttatttatttatttatttatttatttatttatttatttatttattt 2820

```

```

Db 2761 ttccattttatttatttatttatttatttatttatttatttatttatttatttatttattt 2820
Qy 2821 ctctgttttaagaatatttaaaaaacattgtatcatatagatgcttgaagact 2880
Db 2821 ctctgttttaagaatatttaaaaaacattgtatcatatagatgcttgaagact 2880
Qy 2881 agcttctactacactacaagataatgattctcatgtatgctcatalaaacctgcagatgat 2940
Db 2881 agcttctactacactacaagataatgattctcatgtatgctcatalaaacctgcagatgat 2940
Qy 2941 ttccagagtgctcagtaactgttaattacaatcccatgaagctgtaaaagaatgacctac 3000
Db 2941 ttccagagtgctcagtaactgttaattacaatcccatgaagctgtaaaagaatgacctac 3000
Qy 3001 gttctgtatacagctgtgtgtgtttgtatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3060
Db 3001 gttctgtatacagctgtgtgtgtttgtatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3060
Qy 3061 ctgagctctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3120
Db 3061 ctgagctctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3120
Qy 3121 tccatcattgttttaagctgtgaattttctcccaatggaatgtgaatgaatgaatgaatgt 3180
Db 3121 tccatcattgttttaagctgtgaattttctcccaatggaatgtgaatgaatgaatgaatgt 3180
Qy 3181 ttgtcatcaataatgttaataataaaaaaa 3217
Db 3181 ttgtcatcaataatgttaataactaaaaaa 3217

RESULT 2
X90786
ID X90786 standard; DNA; 3190 BP.
XX
AC X90786;
XX
DT 13-JAN-2000 (first entry)
XX
DE DNA encoding human stauflen protein.
XX
KW Human stauflen protein; hstau; protein lysate; lung; kidney; testis;
KW ovary; brain; heart; telomerase RNA; complex; catalytic subunit;
KW telomerase assembly; maturation; transport; regulation; interaction;
KW inhibitory agent; antibody; immunoassay; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 345..1835
FT FT /*tag= a
FT FT /product= "Human stauflen protein"
FT FT /note= "binds to telomerase RNA"
FT FT /*tag= b
XX
PN MO9951255-A1.
XX
PD 14-OCT-1999.
XX
PF 06-APR-1999; 99MO-US07533.
XX
PR 06-APR-1998; 98US-0080783.
XX
PI Greider CW, Le S;
XX
DR WPI: 1999-620168/53.
XX
PT P-PSDB: Y26335.
XX
PT Human stauflen polypeptide useful in methods for identifying telomerase
PT inhibitors -

```

XX Claim 13: Page 41-42: 50pp: English.

CC The present sequence is an isolated DNA encoding human strafen (hstraf)  
CC protein. The hstraf protein was detected in protein lysates from lung,  
CC kidney, testis and ovary, but not in brain and heart. It binds  
CC specifically to human telomerase RNA and can complex with the  
CC telomerase catalytic subunit. The hstraf protein may play a role in  
CC telomerase assembly, maturation, transport and regulation. Interaction  
CC between hstraf and telomerase can be inhibited by administering inhibitory  
CC agents to the cell. These telomerase inhibitors can be identified using  
CC hstraf protein. Antibodies generated against hstraf can be used in various  
CC immunoassays.

XX Sequence 3190 BP: 880 A: 788 C: 736 G: 786 T: 0 other:

Query Match 97.3%; Score 3131.4; DB 20; Length 3190;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 3146; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

OY 56 cgcgtctctctcgtccgcctccttcttgacgcgcctcccccccgccggcgccgc 115  
DB 38 cgcgtccctctcgtccgcctccttcttgacgcgcctcccccccgccggcgccgc 97  
OY 116 gctctctcacgagccacgcgcctcttccctccttgctccttcttctcctctttt 175  
DB 98 gctctctcacgagccacgcgcctccttccctccttgctccttcttctcctctttt 157  
OY 176 tctctctctctcctcctcctcgcgcacgcgcgcagagccgcggcgggagagctc 235  
DB 158 tctctctctctcctcctcctcgcgcacgcgcgcagagccgcggcgggagagctc 217  
OY 226 ggagagagagccagaggttatataaccacttaacctcagaactyacaagaacacat 295  
DB 218 ggagagagagccagaggttatataaccacttaacctcagaactyacaagaacacat 277  
OY 296 gtctcttgaaacgcctctttttaaaaaagaagacataccctactgtagaactaatagc 355  
DB 278 gtctcttgaaacgcctctttttaaaaaagaagacataccctactgtagaactaatagc 337  
OY 356 actgtgacatgaacttggaaaaaaacaaatgtataagcctgttgaccccttactcgcgat 415  
DB 338 actgtgacatgaacttggaaaaaaacaaatgtataagcctgttgaccccttactcgcgat 397  
OY 416 gcaatccacactataactatacaacaatgagagaggtgctatacccccgaagttacttacc 475  
DB 398 gcaatccacactataactatacaacaatgagagaggtgctatacccccgaagttacttacc 457  
OY 476 attccagttccacacttacttacttacttaagaatgtgaactttctgtggagagacagaatttaa 535  
DB 458 attccagttccacacttacttacttacttacttaagaatgtgaactttctgtggagagacagaatttaa 517  
OY 536 tggcaaaagaaagaacagaacagatgctgcgaacacagatgctgctgcgaagcgtttaagat 595  
DB 518 tggcaaaagaaagaacagaacagatgctgcgaacacagatgctgctgctgcgaagcgtttaagat 577  
OY 556 cctgcagagaatggccccctgcagagagagctgagaggttgaaatggaagaagaatccgaaaga 655  
DB 578 cctgcagagaatggccccctgcagagagagctgagaggttgaaatggaagaagaatccgaaaga 637  
OY 656 aaatctcaataaacttgaaataagtcgaagtgtttgagattggaacttaagcgaacttttcc 715  
DB 638 aaatctcaataaacttgaaataagtcgaagtgtttgagattggaacttaagcgaacttttcc 697  
OY 716 tgtgaatttcgagtgagcgcggagagatgagccacccacacatgaagaactttgtgaccaa 775  
DB 698 tgtgaatttcgagtgagcgcggagagatgagccacccacacatgaagaactttgtgaccaa 757  
OY 776 ggttcggtctgggagagcttctgtgggagaggtgaaagggaagaagaattccaagaa 835  
DB 758 ggttcggtctgggagagcttctgtgggagaggtgaaagggaagaagaagaattccaagaa 817

OY 836 aaatgcgccacatagctgtctcttgaggagctgagaagatctaccgcccctgcctcagttga 895  
DB 818 aaatgcgccacatagctgtctcttgaggagctgagaagatctaccgcccctgcctcagttga 877  
OY 896 accgagtaaaagcctagaatcaaaagaaacaaaccacatgtccaagccacagacagcc 955  
DB 878 accgagtaaaagcctagaatcaaaagaaacaaaccacatgtccaagccacagacagcc 937  
OY 956 agaataatgcaggaggaatcaatccgatttagccgacttgcccgagatccagcagaacaaa 1015  
DB 938 agaataatgcaggaggaatcaatccgatttagccgacttgcccgagatccagcagaacaaa 997  
OY 1016 ggagaagagccagagatcacagctctctcacagagagagcctcccgcccgagagaggt 1075  
DB 998 ggagaagagccagagatcacagctctctcacagagagagcctcccgcccgagagaggt 1057  
OY 1076 tgtatgcaggttgaaaggttgaaacacacatgcagaaaggaacgagccacacagaaggt 1135  
DB 1058 tgtatgcaggttgaaaggttgaaacacacatgcagaaaggaacgagccacacagaaggt 1117  
OY 1136 ggcacaagcgaatgtcagccaggaacatgtctgagatccttggttccaagtccgacgc 1195  
DB 1118 ggcacaagcgaatgtcagccaggaacatgtctgagatccttggttccaagtccgacgc 1177  
OY 1196 gcaagccacaaacccgcacactcaagatcagagagagacacacacataagaacacaggg 1255  
DB 1178 gcaagccacaaacccgcacactcaagatcagagagagacacacacataagaacacaggg 1237  
OY 1256 tggaaagaaagttaaccttttttgaaacttgctctgggagatgaaatgggactagtaata 1315  
DB 1238 tggaaagaaagttaaccttttttgaaacttgctctgggagatgaaatgggactagtaata 1297  
OY 1316 agagatgagttcagaatgagccttatctaagtatcataagcagcgtctgctgagatcttcc 1375  
DB 1298 agagatgagttcagaatgagccttatctaagtatcataagcagcgtctgctgagatcttcc 1357  
OY 1376 catgtgcggagagtcgcgcacagcgtctgagaggttgaaagagacatcacacaaagat 1435  
DB 1358 catgtgcggagagtcgcgcacagcgtctgagaggttgaaagagacatcacacaaagat 1417  
OY 1436 taacagagcagctccgaatcctctgcgaagccacggttaactgcattgataagccgagat 1495  
DB 1418 taacagagcagctccgaatcctctgcgaagccacggttaactgcattgataagccgagat 1477  
OY 1496 gtgtatgtgggacactctgcgcacagccgagacacattttaagaatacatctcttcgg 1555  
DB 1478 gtgtatgtgggacactctgcgcacagccgagacacattttaagaatacatctcttcgg 1537  
OY 1556 ccaagtaacccaatgagacactctacagacccctctgaagcaactgagactatcttccagat 1615  
DB 1538 ccaagtaacccaatgagacactctacagacccctctgaagcaactgagactatcttccagat 1597  
OY 1616 ccaaggtatccaggttgaaatacaagaacttcccaaaaaacaaagaagaatttgcatac 1675  
DB 1598 ccaaggtatccaggttgaaatacaagaacttcccaaaaaacaaagaagaatttgcatac 1657  
OY 1676 tcttataatgtctctctcagccacccctctgatacagcatgtatccgcaagagatgta 1735  
DB 1658 tcttataatgtctctctcagccacccctctgatacagcatgtatccgcaagagatgta 1717  
OY 1736 gtctgcacatgatatgagctgcgctgaacacacttaagttgcctgagcttgagtgagaca 1795  
DB 1718 gtctgcacatgatatgagctgcgctgaacacacttaagttgcctgagcttgagtgagaca 1777  
OY 1796 aagtaacagatgccaagaacacagaaacgacaaatgctctgtctggagaggtgctgaac 1855  
DB 1778 aagtaacagatgccaagaacacagaaacgacaaatgctctgtctggagaggtgctgaac 1837  
OY 1856 ctttcttgccatgaaacattataaatacccaacataatacttgaaataactgaaactggc 1915  
DB 1838 ctttcttgccatgaaacattataaatacccaacataatacttgaaataactgaaactggc 1897  
OY 1916 ttgaaatcttgaaatctctgatacctccagtgaggccgagagacacggttggtlaaagat 1975

```

Db      1898  ttgaaatttggaatttcgataccctcagtgccgagagacagcgtgtaagaat 1957
      1976  gtggagcagcagcagggaagaagaacaagaagagcgctgtggcc-ggctgaact 2034
      1958  gtggagcagcagcagggaagaagaacaagaagagcgctgtggccggtgtgact 2017
      2035  gtgctggggttctgtgtagcgaactcgttgacctgcgtccctcgcgaatgcagct 2094
      2018  gtgctggggttctgtgtagcgaactcgttgacctgcgtccctcgcgaatgcagct 2077
      2095  gctgtggggaagaagagcgctgcgcagcagctgtctcccgagacacagcagatccac 2154
      2078  gctgtggggaagaagagcgctgcgcagcagctgtctcccgagacacagcagatccac 2137
      2155  accctggagcagcctcgcgtgtgtgttctttccctgtgtgaagaagaagcagcag 2214
      2138  accctggagcagcctcgcgtgtgtgttctttccctgtgtgaagaagaagcagcag 2197
      2215  acccttctcaagctgtgctcactcactcagacatlgagacaacccctgagacagcag 2274
      2198  acccttctcaagctgtgctcactcactcagacatlgagacaacccctgagacagcag 2257
      2275  agagagcgcttggaccgccccagagctaaagcagcagagaagaatcgaatgtctctac 2334
      2258  agagagcgcttggaccgccccagagctaaagcagcagagaagaatcgaatgtctctac 2317
      2335  tcaagcgtagaccaactttctagtgtgtccagcgccccacacccctcgcagtaaccac 2394
      2318  tcaagcgtagaccaactttctagtgtgtccagcgccccacacccctcgcagtaaccac 2377
      2395  caccacacgcttctctcccaacagtgatcgtatcttagttcatcttcttctt 2454
      2378  caccacacgcttctctcccaacagtgatcgtatcttagttcatcttcttctt 2437
      2455  gattgatacactataataaatttcatcttgagaattctcgaatgtatcgtatga 2514
      2438  gattgatacactataataaatttcatcttgagaattctcgaatgtatcgtatga 2497
      2515  tagcacagtttgaaactgtctgagagctgaacttatacaataacacgaagaagatc 2574
      2498  tagcacagtttgaaactgtctgagagctgaacttatacaataacacgaagaagatc 2557
      2575  atatcactgtatgtgttagacatttattcatctgactgaaccagagcagtttcaag 2634
      2558  atatcactgtatgtgttagacatttattcatctgactgaaccagagcagtttcaag 2617
      2635  tgatgcaaatgtgtgcccctcgtgtcagctgaaacagctccctgagcttcaaaaacttg 2694
      2618  tgatgcaaatgtgtgcccctcgtgtcagctgaaacagctccctgagcttcaaaaacttg 2677
      2695  aataagctcccaagctgtgataaattgtgacaatttggaattttaacttaagatc 2754
      2678  aataagctcccaagctgtgataaattgtgacaatttggaattttaacttaagatc 2737
      2755  attggtcatttttatttcaatttattttgttaatgaaacagagacttaaatgac 2814
      2738  attggtcatttttatttcaatttattttgttaatgaaacagagacttaaatgac 2797
      2815  ttgatctcgttttaagaattatlaaaaaacatgtgtatctatacatatgcttga 2874
      2798  ttgatctcgttttaagaattatlaaaaaacatgtgtatctatacatatgcttga 2857
      2875  ggaactagcttcaactcactacagatatgatacctcagatagtcacataaactcag 2934
      2858  ggaactagcttcaactcactacagatatgatacctcagatagtcacataaactcag 2917
      2935  agtaattccagagtgctgatactgttaattatcctcattagggctgaagaagatg 2994
      2918  agtaattccagagtgctgatactgttaattatcctcattagggctgaagaagatg 2977
      2995  acctagcttctgtatatacagctgtgtccttltgatatgtgttactacacagaagt 3054

```

```

Db      2978  acctagcttctgtatatacagctgtgtccttltgatatgtgttactgtacacagaagt 3037
      3055  tgtgactggagcctcgtcgtgtgtgtccgtatggaagaacctgtgacctgtgagttagt 3114
      3038  tgtgacacgagcctcgtcgtgtgtgtccgtatggaagaacctgtgacctgtgagttagt 3097
      3115  actgctccatctcatctgttcttcagctgaattttctcccccgtgaatgtgaataaact 3174
      3098  actgctccatctcatctgttcttcagctgaattttctcccccgtgaatgtgaataaact 3157
      3175  aagtgttgcatacataaataatgtgaataa 3207
      3158  aagtgttgcatacataaataatgtgaataa 3190

RESULT      3
Z93324
ID      Z93324 standard; cDNA; 3142 BP.
AC      Z93324;
XX
XX      04-JUL-2000 (first entry)
DE      Human stauflen cDNA.
XX
XX      Stauflen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KM      HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW      acquired immune deficiency syndrome; protease; human; ss.
XX
XX      Homo sapiens.
XX
XX      Key      Location/Qualifiers
FH      CDS      288..1193
FT      CDS      /*tag= a
FT      FT      /product= Stauflen protein
XX
XX      CA2238656-A1.
PN
XX
XX      22-NOV-1999.
PD
XX
XX      22-MAY-1998; 98CA-2238656.
PF
XX
XX      22-MAY-1998; 98CA-2238656.
PR
XX
XX      (UYMO-) UNIV MONTREAL.
PA
XX      Duchaine T, Wickham L, Cohen EA, Luo M, Desrochers L;
PI      Moulard AJ;
XX
XX      WPI: 2000-246924/22.
DR      P-PSDB; Y83108.
XX
XX      Isolated mammalian or Caenorhabditis elegans stauflen proteins useful
PT      for treating retroviral infections especially human immunodeficiency
PT      virus infections (HIV)
XX
XX      Claim 4; Figure 1b; 96pp; English.
PS
XX
XX      Stauflen is a RNA binding protein which interacts with double stranded
CC      RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC      transport and localization. Mammalian SP contain multiple ds
CC      RNA-binding domains. Chimeric proteins comprising portions of
CC      stauflen with proteins with RNase or protease activity can be used to
CC      treat virus infections. The RNase or protease activity of the fusion
CC      protein prevents proper maturation of the virus. Mammalian stauflen
CC      proteins seem to recognise double stranded RNA structure rather than
CC      any sequence specific position. 2 bacterially-expressed fusion
CC      proteins used in an RNA-binding assay (his/hsp (human SP) and
CC      MBP/hsp (murine SP) both proteins strongly bound double stranded
CC      RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC      RNA/DNA hybrids. Stauflen proteins may therefore be useful in the
CC      treatment of RNA virus infections, especially those caused by
CC      retroviruses, in particular human immunodeficiency virus. This

```

CC sequence is an alternatively spliced sequence to the one given in  
CC Z93322.  
XX  
SQ Sequence 3142 BP: 862 A; 778 C; 733 G; 769 T; 0 other;

Query Match 94.9%; Score 3053.8; DB 21; Length 3142;  
Best Local Similarity 97.6%; Pred. No. 0;  
Matches 3140; Conservative 0; Mismatches 2; Indels 75; Gaps 1;

QY 1 actctctcgccggctgctggcgccctgagcgctcttcagcgcttctgogcgcgctgcgct 60  
DB 1 actctctcgccggctgctggcgccctgagcgctcttcagcgcttctgogcgcgctgcgct 60  
QY 61 ctctctcgccgtcccgcttcttcttgaccgctctcccccggcgccggcgccgcctc 120  
DB 61 ctctctcgccgtcccgcttcttcttgaccgctctcccccggcgccggcgccgcctc 120  
QY 121 ctccacggcgactccgctctctccctctccctctgcctctctctctctctctcttctc 180  
DB 121 ctccacggcgactccgctctctccctctccctctgcctctctctctctctctcttctc 180  
QY 181 ctctctctccctctctcccgccacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 240  
DB 181 ctctctctccctctctcccgccacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 240  
QY 241 agcagccgaggttatcaaccaacttaactctcagaaacgaaacaaactgttcc 300  
DB 241 agcagcc----- 247  
QY 301 tggaaagccctctttaaagaagaacataaccctactgttagaactaaatgcactgt 360  
DB 248 -----agaaagcataccctactgttagaactaaatgcactgt 285  
QY 361 gcatgaacttggaaaaaacaatgataagcgttgcacctactctcggaatgcagt 420  
DB 286 gatatgaacttggaaaaaacaatgataagcgttgcacctactctcggaatgcagt 345  
QY 421 ccacctatactacaacatgagagaggtgtctatcccccggagttactttccacttc 480  
DB 346 ccacctatactacaacatgagagaggtgtctatcccccggagttactttccacttc 405  
QY 481 cagttccaccttactttaaacttcaagtggaaacttctgtggagagacagcaatttaatgca 540  
DB 406 cagttccaccttactttaaacttcaagtggaaacttctgtggagagacagcaatttaatgca 465  
QY 541 aaggaaagacagacagcgtctgcgaacaacgatagtcgtgcacaaagcgttggagatctctgc 600  
DB 466 aaggaaagacagacagcgtctgcgaacaacgatagtcgtgcacaaagcgttggagatctctgc 525  
QY 601 aagaatgagccctgcagagagagctgaggttgaatggaagagaatccgaaagaaatc 660  
DB 526 aagaatgagccctgcagagagagctgaggttgaatggaagagaatccgaaagaaatc 585  
QY 661 tcaataaatctgaaataaactcaagtgttgcagatgcactttaaaccgaaacttgcctgtga 720  
DB 586 tcaataaatctgaaataaactcaagtgttgcagatgcactttaaaccgaaacttgcctgtga 645  
QY 721 atttcgaggttgcgcgggagagatgccccacacatgaagaacttctgcaccaagttt 780  
DB 646 atttcgaggttgcgcgggagagatgccccacacatgaagaacttctgcaccaagttt 705  
QY 781 cagtttggagagtttgcgggagaggtgaaagggaaggaagaagatttcaagaanaaatg 840  
DB 706 cagtttggagagtttgcgggagaggtgaaagggaaggaagaagatttcaagaanaaatg 765  
QY 841 ccgcacatactgttctctggagagctgaaagaagttaaccgcccctgcctgtaagtgaacgag 900  
DB 766 ccgcacatactgttctctggagagctgaaagaagttaaccgcccctgcctgtaagtgaacgag 825  
QY 901 taaagccttagaatcaaaaaaagaacaaacacatagtcagccacagacagaccagaat 960  
DB 826 taaagccttagaatcaaaaaaagaacaaacacatagtcagccacagacagaccagaat 885

QY 961 atgcccagggagatcaatccgatttagcgactgcccagatccacgacgacaaaaagagaa 1020  
DB 886 atgcccagggagatcaatccgatttagcgactgcccagatccacgacgacaaaaagagaa 945  
QY 1021 aggaagcagagtagacagctctctcaacagagcgagcctcccgcgccgagggagttgtga 1080  
DB 946 aggaagcagagtagacagctctctcaacagagcgagcctcccgcgccgagggagttgtga 1005  
QY 1081 tgcaggttgaagtttggaaaccacactgcgcagaaaggaaacgggcacccaagaaggtgtgcca 1140  
DB 1006 tgcaggttgaagtttggaaaccacactgcgcagaaaggaaacgggcacccaagaaggtgtgcca 1065  
QY 1141 agcgacatgcagccggagaaatgctgagatccttggttttcaaatgcccgacggcgacgc 1200  
DB 1066 agcgacatgcagccggagaaatgctgagatccttggttttcaaatgcccgacggcgacgc 1125  
QY 1201 ccaccaaacccgcactcaatgacagagagaaagacaccaaaagaaacagggatgagaa 1260  
DB 1126 ccaccaaacccgcactcaatgacagagagaaagacaccaaaagaaacagggatgagaa 1185  
QY 1261 gaaaagttaaccttttttgaaacttgctctggtggatgtaaaatggtgacttaataaagag 1320  
DB 1186 gaaaagttaaccttttttgaaacttgctctggtggatgtaaaatggtgacttaataaagag 1245  
QY 1321 atgagttcagagatgccttaacttaagtcacacagcgtgcctctggaatcttcccatg 1380  
DB 1246 atgagttcagagatgccttaacttaagtcacacagcgtgcctctggaatcttcccatg 1305  
QY 1381 tgcgcagagtcgcgccagcgtgtagaggttagtcaagaagacataccacaaagatttacc 1440  
DB 1306 tgcgcagagtcgcgccagcgtgtagaggttagtcaagaagacataccacaaagatttacc 1365  
QY 1441 gggcgagcttcggaatcctctgccaagggccacggtactgcattgtagccgaagttgtgt 1500  
DB 1366 gggcgagcttcggaatcctctgccaagggccacggtactgcattgtagccgaagttgtgt 1425  
QY 1501 atgggagccacttcgcccacagccgagacacattttaaagataacatctcttcagcgacg 1560  
DB 1426 atgggagccacttcgcccacagccgagacacattttaaagataacatctcttcagcgacg 1485  
QY 1561 taaccatagacctctcagagacacctctgagaaactgtgactatcttccagaatccag 1620  
DB 1486 taaccatagacctctcagagacacctctgagaaactgtgactatcttccagaatccag 1545  
QY 1621 gattccaggttgaatacaagaacttcccaaaaacaagaagcaattgtatctctta 1680  
DB 1546 gattccaggttgaatacaagaacttcccaaaaacaagaagcaattgtatctctta 1605  
QY 1681 tcaatgtcctctcagccactctgatacagcatgtatactggcaagatgtgtgaatct 1740  
DB 1606 tcaatgtcctctcagccactctgatacagcatgtatactggcaagatgtgtgaatct 1665  
QY 1741 gccatgatactgctgcgtcgtgaacatcttaagttgtcgtctgagtttggaaccaagaatga 1800  
DB 1666 gccatgatactgctgcgtcgtgaacatcttaagttgtcgtctgagtttggaaccaagaatga 1725  
QY 1801 cagagatgccaagaacaggaacagcaactgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1860  
DB 1726 cagagatgccaagaacaggaacagcaactgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1785  
QY 1861 ctggcatalaacatataataatcccaacatataatactgaaataactgaaactgccttga 1920  
DB 1786 ctggcatalaacatataataatcccaacatataatactgaaataactgaaactgccttga 1845  
QY 1921 aaatttgaattctgtatacctccagttgggcccagagacacggttggtaagaatgtggg 1980  
DB 1846 aaatttgaattctgtatacctccagttgggcccagagacacggttggtaagaatgtggg 1905  
QY 1981 cagcgacagggagaaagaaacaaagaaacaaagggcggtcgtgtggcggtgcgtgtgtgtgt 2040  
DB 1906 cagcgacagggagaaagaaacaaagaaacaaagggcggtcgtgtggcggtgcgtgtgtgtgtgt 1965

QY	2041	ggtttgttgtgtatgtgtccactctgtgtgactctgtgtcttctaagcaataagactgtcctgt	2100
Db	1966	gggtttgtgtgtatgtatgtgcacactctgtgtactctgtgtcttctctactacgaataagactgtcctgt	2025
QY	2101	ggggagaagaaggctgtcccccagcagctgtgtcttccccgggagacacagatgtccacacctgt	2160
Db	2026	ggggagaagaaggctgtcccccagcagctgtgtcttccccgggagacacagatgtccacacccctgt	2085
QY	2161	ggcacctccgtgttctgtctcttcttcttccctctgtgtgtgaagaagaacggtacagaccct	2220
Db	2086	ggcacctccgtgttctgtctcttcttcttcttcttccctctgtgtgtgaagaagaacggtacagaccct	2145
QY	2221	tctcaagctgtgtcactacacgaacatgtgtgagacaaacccctgtgacgcacatgtccagatagag	2280
Db	2146	tctcaagctgtgtcactacacgaacatgtgtgagaaacccctgtgacgcacatgtccagatagag	2205
QY	2281	gaccttgcacggcccccagagctcaaaagacccagagaaatacaaatgtcttccactacagc	2340
Db	2206	gaccttgcacggcccccagagctcaaaagacccagagaaatacaaatgtcttccactacagc	2265
QY	2341	tgcaccaacttcttctagtgtgtccacagcccccacacacactcctctgtgagtacacacacacac	2400
Db	2266	tgcaccaacttcttctagtgtgtccacagcccccacacacactcctctgtgagtacacacacacac	2325
QY	2401	cactgtcttctcttcccaacagatgactctgtatcttctagtgttccatatttctctttgtatga	2460
Db	2326	cactgtcttctcttcccaacagatgactctgtatcttctagtgttccatatttctctttgtatga	2385
QY	2461	tatgcacactataaaatttccaatttgatggaatttcccaattgtatctatagttaaatagcac	2520
Db	2386	tatgcacactataaaatttccaatttgatggaatttcccaattgtatctatagttaaatagcac	2445
QY	2521	agtttgcgaactgttctgtgagactgtacttatacaatactaaacggacaagaatcatatcc	2580
Db	2446	agtttgcgaactgttctgtgagactgtacttatacaatactaaacggacaagaatcatatcc	2505
QY	2581	atgtgtatgtgtgtatagaacatttatttccatttgcactaacccagagactttcacgtatgc	2640
Db	2506	atgtgtatgtgtgtatagaacatttatttccatttgcactaacccagagactttcacgtatgc	2565
QY	2641	aaattgtgtgacctctgtgttcaagctgtgaaaacagtccttgagcttcaaaaaccttgataag	2700
Db	2566	aaattgtgtgacctctgtgttcaagctgtgaaaacagtccttgagcttcaaaaaccttgataag	2625
QY	2701	tctccacacggtgtataaaatttgagacaaatttaaggaattttaaactttagatgacatcttg	2760
Db	2626	tctccacacggtgtataaaatttgagacaaatttaaggaattttaaactttagatgacatcttg	2685
QY	2761	tcccatatttatttcaatttatttcttgtttaaalycaaacagagacttaaalgtaacattgtat	2820
Db	2686	tcccatatttatttcaatttatttcttgtttaaalycaaacagagacttaaalgtaacattgtat	2745
QY	2821	cctctgtttttaaagaattataaaaaacatgtgtatctatacatatagctctcttgaagact	2880
Db	2746	cctctgtttttaaagaattataaaaaacatgtgtatctatacatatagctctcttgaagact	2805
QY	2881	agccttccactaacactaacagagatagtatcccatgagtatgcataataaacctgcagagat	2940
Db	2806	agccttccactaacactaacagagatagtatcccatgagtatgcataataaacctgcagagat	2865
QY	2941	tctccagagtgctcgtatatactgtttaattatactcccatctaaagtgtaaaaagatgtaccctac	3000
Db	2866	tctccagagtgctcgtatatactgtttaattatactcccatctaaagtgtaaaaagatgtaccctac	2925
QY	3001	gtttctgtatacagctgtgtgtcttcttgaatgtgtgttactgtatacacagaagtgtgtgca	3060
Db	2926	gtttctgtatacagctgtgtgtcttcttgaatgtgtgttactgtatacacagaagtgtgtgca	2985
QY	3061	ctgaagacctctgt	3120
Db	2986	ctgaagacctctgt	3045
QY	3121	tccattccattgttttaacgtgtgaaatttcttcccatctgaaatgtlaagttaaacttaagtgt	3180

Db 3046 tccatccatcgttllaacgctggaattcttcctcccacggaaatgtaagtcaaaactaaagtgt 3105

Oy 3181 ttgtcatcaatlaaagtgttaactactaaaaa 3217

Db 3106 ttgtcatcaatlaaagtgttaactactaaaaa 3142

RESULT	4
293327	
ID	293327 standard; cDNA; 3260 BP.

AC Z93327;

DT 04-JUL-2000 (first entry)

DE Human staußen cDNA

KW Staufen; RNA binding

KW acquired immu

OS Homo sapiens.

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

ET / \*tag= a

XX

XX

XX

XX

XX

XX

PI Mouland AJ;

DR WPI; 2000-240

XX

PT for treating

XX

XX

CC RNA and/or rRNA

CC RNA-binding o

CC treat virus

proteins seen

CC proteins used

CC RNA. Both fus

CC treatment of

CC sequence is a

XX

Query Match

Query Match

94.48; Score 3037.4; DB 21; Length 3260;



Best Local Similarity 96.6%; Pred. No. 0;  
Matches 3153; Conservative 0; Mismatches 61; Indels 49; Gaps 3;

```
OY 1 actctctgcgggctctgcggcgctctgaagcgtcttcaagcgtttgcggcgcgctgcgt 60
Db 1 actctctgcgggctctgcggcgctctgaagcgtcttcaagcgtttgcggcgcgctgcgt 60
OY 61 ctctctgcgctcccgcttctcttgaacgcgtcccccgcggcgcgcgccgctc 120
Db 61 ctctctgcgctcccgcttctcttgaacgcgtcccccgcggcgcgcgccgctc 120
OY 121 ctccacgacctcgctcttctccctcgtlccc----- 157
Db 121 ctccacgacctcgctcttctccctcgtlccc----- 157
OY 158 -ttctccctctcttctcttctcttctcttctcttctcttctcttctcttctctt 216
Db 181 ttgcgctctgtccgcccagctgagctgagctgagctgagctgagctgagctgagct 240
OY 217 ccgcccggggga-----cgagctcgagagcagcagcagagtt 254
Db 241 cctcccaagtcagcagatttcccaactcagccctcccgataagctgagattacagagtt 300
OY 255 attaacaccttaacctctcaagaactgacaagaacaacatgttctctggaacgctctt 314
Db 301 attaacaccttaacctctcaagaactgacaagaacaacatgttctctggaacgctctt 360
OY 315 tttaaaaagaagaagcaaaacccctactctgtaagaactaagcactgagcactgaaacttga 374
Db 361 tttaaaaagaagaagcaaaacccctactctgtaagaactaagcactgagcactgaaacttga 420
OY 375 aaaaaaaccaatgtatagaactgttgaaccttaactcgcgagtcagtcacacctaatactac 434
Db 421 aaaaaaaccaatgtatagaactgttgaaccttaactcgcgagtcagtcacacctaatactac 477
OY 435 aacaatgagagagtgcttaaccoccgaggtactttaccocatltccacttccacttta 494
Db 478 aacaatgagagagtgcttaaccoccgaggtactttaccocatltccacttccactttaa 537
OY 495 ctttatacaagtgaactcttctgtgtggagagacagcaatttaatgagcaagaagaagaaga 554
Db 538 ctttatacaagtgaactcttctgtgtggagagacagcaatttaatgagcaagaagaagaaga 597
OY 555 cagcgtctgcaaaacagatctgtctgcacaagcgttgaagatccctgagaatgagccctct 614
Db 598 cagcgtctgcaaaacagatctgtctgcacaagcgttgaagatccctgagaatgagccctct 657
OY 615 ccaagagaagctggaagtgaaatggaagaagaatccgaagaagaagaatccaaataacttga 674
Db 658 ccaagagaagctggaagtgaaatggaagaagaatccgaagaagaagaatccaaataacttga 717
OY 675 atcaagtcaagttcttgagattgcaacttaacggaacttccctgtgaattctgaagtgagcc 734
Db 718 atcaagtcaagttcttgagattgcaacttaacggaacttccctgtgaattctgaagtgagcc 777
OY 735 cgggagagtggtcccaaccacatgaagaactctgtgaaccaagttccggtctggggagttt 794
Db 778 cgggagagtggtcccaaccacatgaagaactctgtgaaccaagttccggtctggggagttt 837
OY 795 gtggggggaagtggaagggaaagcaagaagattcaagaaaaaatgacgcaatagctgtt 854
Db 838 gtggggggaagtggaagggaaagcaagaagattcaagaaaaaatgacgcaatagctgtt 897
OY 855 ctltggagagcttgaagaagttaccgcccctgcctgtcaagttgaaacggttaagcttgaatc 914
Db 898 ctltggagagcttgaagaagttaccgcccctgcctgtcaagttgaaacggttgaatc 957
OY 915 aaaaagaacaacaacccatagtcacagccacagacagaccagatatgtgcagaggtatc 974
Db 958 aaaaagaacaacaacccatagtcacagccacagacagaccagatatgtgcagaggtatc 1017
OY 975 aatccgattagcgaactgtgccagatccagcagcgaaaaaagagagagcagaatgac 1034
Db 1017 aatccgattagcgaactgtgccagatccagcagcgaaaaaagagagagcagaatgac 1034
```

```
Db 1018 aatccgattagcgaactgtgccagatccagcagcgaaaaaagagagagcagaatgac 1077
OY 1035 aagctctcaacagagcgaagcctcccgccgacgagagagttgtgtatgagtggaagtt 1094
Db 1078 aagctctcaacagagcgaagcctcccgccgacgagagagttgtgtatgagtggaagtt 1137
OY 1095 ggaacaacacatctgcaagaagaaacggtcaccaacaagaagtggtccaaagcgaatgacgc 1154
Db 1138 ggaacaacacatctgcaagaagaaacggtcaccaacaagaagtggtccaaagcgaatgacgc 1197
OY 1155 ggaacaatctgtgagatctctgtgtttcaaatgctccgacgacgacccaacccgca 1214
Db 1198 ggaacaatctgtgagatctctgtgtttcaaatgctccgacgacgacccaacccgca 1257
OY 1215 ctcaagtctgagagagagacacccaataaagaacgggagatggaagaagttaacctt 1274
Db 1258 ctcaagtctgagagagagacacccaataaagaacgggagatggaagaagttaacctt 1317
OY 1275 ttgtaacctgtgctgtggtgagaaatgtggaactagtaataaagagagatgactcaagatg 1334
Db 1318 ttgtaacctgtgctgtggtgagaaatgtggaactagtaataaagagagatgactcaagatg 1377
OY 1335 ccttatctaagtcatcagcagctgctgtctggaattcttccatggtgcccgaggtgcgc 1394
Db 1378 ccttatctaagtcatcagcagctgctgtctggaattcttccatggtgcccgaggtgcgc 1437
OY 1395 cagcgtgtagaggttagttaaagacatcaccaagaagattttaaaggcgaactccgaat 1454
Db 1438 cagcgtgtagaggttagttaaagacatcaccaagaagattttaaaggcgaactccgaat 1497
OY 1455 cctgtccaaagccacagctaaactgcaatgaaagcggagctgtgtatgtatgtgggacactgc 1514
Db 1498 cctgtccaaagccacagctaaactgcaatgaaagcggagctgtgtatgtatgtgggacactgc 1557
OY 1515 cccaacgcccgaagccaattttaaagaataacatctctcagggccaagtaccccaatgacct 1574
Db 1558 cccaacgcccgaagccaattttaaagaataacatctctcagggccaagtaccccaatgacct 1617
OY 1575 ctcaagagaacctctgtagaactgtgactatcttccagaagtcgaaggattccaggttga 1634
Db 1618 ctcaagagaacctctgtagaactgtgactatcttccagaagtcgaaggattccaggttga 1677
OY 1635 taacaagacttccccaacaacaacaagaagcaattgtatctcttcaacttgcctcctc 1694
Db 1678 taacaagacttccccaacaacaacaagaagcaattgtatctcttcaacttgcctcctc 1737
OY 1695 cagccaactctgtatcagcgaatgtgtatcgtgcagaagatgtgtgactcctgcataatgact 1754
Db 1738 cagccaactctgtatcagcgaatgtgtatcgtgcagaagatgtgtgactcctgcataatgact 1797
OY 1755 ggcgtgaacatcttaagtgtctgtctgtgagttggaaccaagaagtacaagaatgacaaga 1814
Db 1798 ggcgtgaacatcttaagtgtctgtctgtgagttggaaccaagaagtacaagaatgacaaga 1857
OY 1815 acaagaaaacggaacaaatgtctgtgtgtggaggtgtcgtgaacacttctctgccaatgaacca 1874
Db 1858 acaagaaaacggaacaaatgtctgtgtgtggaggtgtcgtgaacacttctctgccaatgaacca 1917
OY 1875 ttataaatcccaacatatatactgtaaaatctactgaactgtcttgaataatcttgaatttc 1934
Db 1918 ttataaatcccaacatatatactgtaaaatctactgaactgtcttgaataatcttgaatttc 1977
OY 1935 tgataccttccagtggtgcgagagacagctgtgttaaagatgtgtggcagcagaaggaag 1994
Db 1978 tgataccttccagtggtgcgagagacagctgtgttaaagatgtgtggcagcagaaggaag 2037
OY 1995 acaacagaacaacaagagagcggtctgtgcgcgtgcagctgtgcgtgtgtgtgtat 2054
Db 2038 acaacagaacaacaagagagcggtctgtgcgcgtgcagctgtgcgtgtgtgtgtat 2097
OY 2055 ggcacactgtgtacctgtgcgtgtccctacgcaatagcagctgtcgtgtgtgtgtgtgtgt 2114
Db 2098 ggcacactgtgtacctgtgcgtgtccctacgcaatagcagctgtcgtgtgtgtgtgtgtgt 2157
```



OY	2115	gcacagcaacgtggtctctccggagacaacagacagatccacacccgggcaactccgtgt	2174
Db	2158	gccacagccagctggtctctcccgggacaacagacagatccaaccccgggcaactccgtgt	2217
OY	2175	tgcctcttttcccccgtgtbaaagaagaaacggtcaagacccctctcaagctggtc	2234
Db	2218	tgcctcttttcccccgtgtbaaagaagaaacggtcaagacccctctcaagctggtc	2277
OY	2235	actcagacacatctgggacaacaccctggagacgatccagagagagcctcttgacgggc	2294
Db	2278	actcagacacatctgggacaacaccctggagacgatccagagagagcctcttgacgggc	2337
OY	2295	ccagagactaaagaacacacagagaaatccaaatgcttctctactagctgacccaactttc	2354
Db	2338	ccagagactaaagaacacacagagaaatccaaatgcttctctactagctgacccaactttc	2397
OY	2355	tagtgtgcacagggccccaacacctctctgagtaaccaacacctaacagctgtctctt	2414
Db	2398	tagtgtgcacagggccccaacacctctctgagtaaccaacacctaacagctgtctctt	2457
OY	2415	ccaacagctgactcgtactcttaatttcattatcttcttgaatgtatatacaactata	2474
Db	2458	ccaacagctgactcgtactcttaatttcattatcttcttgaatgtatatacaactata	2517
OY	2475	aaatttcatttgagaattctccaaitgtatctagttaaataagacaagcttgnaaactg	2534
Db	2518	aaatttcatttgagaattctccaaitgtatctagttaaataagacaagcttgnaaactg	2577
OY	2535	ctgtagcgtgactttccaataatctaaacccgacaaagaatccatccatggtatgtgt	2594
Db	2578	ctgtagcgtgactttccaataatctaaacccgacaaagaatccatccatggtatgtgt	2637
OY	2595	agacattttatctcatcttgactaaccacagagacagtttcagttgcaaatgtgacct	2654
Db	2638	agacattttatctcatcttgactaaccacagagacagtttcagttgcaaatgtgacct	2697
OY	2655	ctgtgtcagctgnaaacagctctgagacttccaanaacctgnaataagctctccacagttg	2714
Db	2698	ctgtgtcagctgnaaacagctctgagacttccaanaacctgnaataagctctccacagttg	2757
OY	2715	ataaatctggacaatttagagaattttaactttagatagatcatcttggtcccatcttatt	2774
Db	2758	ataaatctggacaatttagagaattttaactttagatagatcatcttggtcccatcttatt	2817
OY	2775	cattttattttgtttaaagcaaacaggacttaaatgaaacttgatctctgttttaaga	2834
Db	2818	cattttattttgtttaaagcaaacaggacttaaatgaaacttgatctctgttttaaga	2877
OY	2835	ttatataaaaaacatctgtgtactatacaatagctctcttgagacttagcttcaactaac	2894
Db	2878	ttatataaaaaacatctgtgtactatacaatagctctcttgagacttagcttcaactaac	2937
OY	2895	tacagagatagatcccatctgtagtccatataaaacccgacagatgtatttccagagtctc	2954
Db	2938	tacagagatagatcccatctgtagtccatataaaacccgacagatgtatttccagagtctc	2997
OY	2955	gataccttttaataacatctccatctcattagggctggaanaagaatgaaactcgttttcgtataag	3014
Db	2998	gataccttttaataacatctccatctcattagggctggaanaagaatgaaactcgttttcgtataag	3057
OY	3015	ctgtgtctgcttttgaatgttgttactcgttacaagaagtggtgacgttgaggtcctcggt	3074
Db	3058	ctgtgtctgcttttgaatgttgttactcgttacaagaagtggtgacgttgaggtcctcggt	3117
OY	3075	gttgttccgtatgtgaaacactggttagccctggtcgagtttaagtactggtccatcatctgttt	3134
Db	3118	gttgttccgtatgtgaaacactggttagccctggtcgagtttaagtactggtccatcatctgttt	3177
OY	3135	acgcttggaattttctcccaatgagaatgtgaagaaacacttaagtggttgttaataaaa	3194
Db	3178	acgcttggaattttctcccaatgagaatgtgaagaaacacttaagtggttgttaataaaa	3237

```
Qy 3195 tgytaataactaaaaaaaaa 3217
    |||||
Db 3238 tgytaataactaaaaaaaaa 3260
```

RESULT	5
293322	
ID	293322 standard; cDNA; 3506 BP.

AC 293322;

DT 04-JUL-2000 (first entry)  
XXDE Human S  
yy

KW	Staufen
KW	HTV. hin

KW acquire  
XX

US	Homo sapiens
XX	

FT	key	CDS
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

FT 12

CA22386

PD 22-NOV-

PE 22-MAY-

PR 22-MAY-

PA  
(UYMO-)  
1977

PI Duchain

XX  
DP WPT : 30

DR P-PSDB;  
XX

PT	Isolate for tre
PI	Isolate

XX  
VI  
VICUS I

XX  
PS  
Claim 9

CC RNA and

RNA-binding

CC treat v

protein

protein

CC RNA. BC  
CC RNA/DNA

CC treatment

Sequence	XX	SO
----------	----	----

Best Local Query Match

2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526

—

Db 1 acttcttcgcyggtctgcygagccttgagcgcctcttcacgcgttctgcygagcgtctgcyt 60  
Qy 61 cctctcgcgtcccgcttccctttgacgcgtcccccgcgcgcgcgcgcgcgc 120  
Db 61 cctctcgcgtcccgcttccctttgacgcgtcccccgcgcgcgcgcgcgcgc 120  
Qy 121 ctcacagcgcactcgcctcttccctcctctcgctcccttctctctcttctt 180  
Db 121 ctcacagcgcactcgcctcttccctcctctcgctcccttctctctcttctt 180  
Qy 181 cttctctctccctctctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 240  
Db 181 cttctctctccctctctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 240  
Qy 241 agcagcagagttatctaaccaacttaacctctacagactgaaacaagaacttctc 300  
Db 241 agcagcagagttatctaaccaacttaacctctacagactgaaacaagaacttctc 300  
Qy 301 tggagcgccctctttaaataa----- 322  
Db 301 tggagcgccctctttaaataaagtagaactttagacttcacatgaaatlaacctg 360  
Qy 323 ----- 322  
Db 361 caactgaaagcgttctacgtcatttgctacatttggtgaaagtacatgctcgaagt 420  
Qy 323 ----- 322  
Db 421 caagtgcaggttcaagaaccacatcgtctctctctcagagagccaacttgaaagaac 480  
Qy 323 ----- 322  
Db 481 cagttctctctcagcagccttgatgagtatctctctactactagcctctcgccctct 540  
Qy 323 ----- 322  
Db 541 gaaaatgcaggtagacccaattcaaaactctgttaccctctgcatctatlaaccacc 600  
Qy 323 ----- agaaagcataaccctactgttagaactaaatgacactgtgcatgaactt 371  
Db 601 agtgcagcgtcgcaaaagccttaaccctactgttagaactaaatgacactgtgcatgaactt 660  
Qy 372 ggaataaaaccaaactgtataagcctgttgacacttaactcgcgaatgacgtccactaact 431  
Db 661 ggaataaaaccaaactgtataagcctgttgacacttaactcgcgaatgacgtccactaact 720  
Qy 432 tacaataatgagagagtgcttatacccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 491  
Db 721 tacaataatgagagagtgcttatacccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 780  
Qy 492 ttaacttatcaagtgaactttctgtgtggaagacagcaatttaatgcaagaagaaagaca 551  
Db 781 ttaacttatcaagtgaactttctgtgtggaagacagcaatttaatgcaagaagaaagaca 840  
Qy 552 agacagcgtcgcaaaacacatgctgtctgcaaaagcgttagagaacctgcaagaatgagcc 611  
Db 841 agacagcgtcgcaaaacacatgctgtctgcaaaagcgttagagaacctgcaagaatgagcc 900  
Qy 612 ctgcagcagagcgtcgaggttgaaatggaagaaatcccgaaagaanaatccatlaactc 671  
Db 901 ctgcagcagagcgtcgaggttgaaatggaagaaatcccgaaagaanaatccatlaactc 960  
Qy 672 gaaataagcagaagtgtttgagattgcaactaaacggaacttgctgttgatattcgaggtg 731  
Db 961 gaaataagcagaagtgtttgagattgcaactaaacggaacttgctgttgatattcgaggtg 1020  
Qy 732 gcccgagagagtgccacccacacatgaagaacttgtagccaagtctcggttgaggag 791  
Db 1021 gcccgagagagtgccacccacacatgaagaacttgtagccaagtctcggttgaggag 1080  
Qy 792 ttctgtgaggagagtgaaaggaaagcaagaattccaagaanaaatgcgcgcataagct 851  
Db 1081 ttctgtgaggagagtgaaaggaaagcaagaattccaagaanaaatgcgcgcataagct 1140

Qy 852 gtctctgagagcttgaaagaagttaccgcccctgctctcagttgaaacgagtaaaacctaga 911  
Db 1141 gtctctgagagcttgaaagaagttaccgcccctgctctcagttgaaacgagtaaaacctaga 1200  
Qy 912 atcaaaaagaanaacaaaccatagttcaagccaacagacaagccacagataatgccaaggg 971  
Db 1201 atcaaaaagaanaacaaaccatagttcaagccaacagacaagccacagataatgccaaggg 1260  
Qy 972 atcaatccgattagccagacttgccagatccagccaagaaaagaaggaagccagag 1031  
Db 1261 atcaatccgattagccagacttgccagatccagccaagaaaagaaggaagccagag 1320  
Qy 1032 taacagctctccacagagagagcctcccgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1091  
Db 1321 taacagctctccacagagagagcctcccgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1380  
Qy 1092 gttagaacccaactgacagaaagacggaaccaacaagaaggttgccaagcgaatgca 1151  
Db 1381 gttagaacccaactgacagaaagacggaaccaacaagaaggttgccaagcgaatgca 1440  
Qy 1152 gccgagaaacatgtctgagatccctgttccaagttccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1211  
Db 1441 gccgagaaacatgtctgagatccctgttccaagttccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1500  
Qy 1212 gcaactcaagtcagagagagagacacccaataaagaacccagggatgtgaaagaagaataacc 1271  
Db 1501 gcaactcaagtcagagagagagacacccaataaagaacccagggatgtgaaagaagaataacc 1560  
Qy 1272 ttttttgaaactgctctggtggaatggaataatggactagtaataaagaagatgagttcag 1331  
Db 1561 ttttttgaaactgctctggtggaatggaataatggactagtaataaagaagatgagttcag 1620  
Qy 1332 atgacctatctagatcaagaagcgtcgtcgtgc 1391  
Db 1621 atgacctatctagatcaagaagcgtcgtcgtgc 1680  
Qy 1392 gcccaagcgttagagttagtcagagacatcacaccaaaagatttlaaccaggcagctccg 1451  
Db 1681 gcccaagcgttagagttagtcagagacatcacaccaaaagatttlaaccaggcagctccg 1740  
Qy 1452 aactctgcgaagccaagcgaatgacatgcatagatcccgagagttgttcataggggagcc 1511  
Db 1741 aactctgcgaagccaagcgaatgacatgcatagatcccgagagttgttcataggggagcc 1800  
Qy 1512 tcgccaacagccgagacccatttaagaataacatctcttcgc 1571  
Db 1801 tcgccaacagccgagacccatttaagaataacatctcttcgc 1860  
Qy 1572 cctctcagagacccctctagcaactgacatctctcagagttccagagattccaggtt 1631  
Db 1861 cctctcagagacccctctagcaactgacatctctcagagttccagagattccaggtt 1920  
Qy 1632 gaatacaagacttcccaaaaacaaacaagaacgaattgtatctcttaactatgctcc 1691  
Db 1921 gaatacaagacttcccaaaaacaaacaagaacgaattgtatctcttaactatgctcc 1980  
Qy 1692 tctcagcaccctctgataagcgaatgtatcgcgcgaagatgttgaggtccgcgcatagtatg 1751  
Db 1981 tctcagcaccctctgataagcgaatgtatcgcgcgaagatgttgaggtccgcgcatagtatg 2040  
Qy 1752 gcttcgtgtaaacacttaagtgtgctgtgagttgagccaacaagaatgacagatgcca 1811  
Db 2041 gcttcgtgtaaacacttaagtgtgctgtgagttgagccaacaagaatgacagatgcca 2100  
Qy 1812 agaaacaggaacggaacaaatgtctgtgtgtaggaagttgtagaactttctgccaatgaa 1871  
Db 2101 agaaacaggaacggaacaaatgtctgtgtgtaggaagttgtagaactttctgccaatgaa 2160  
Qy 1872 ccaattataatcccaacataatactgaaataactgaaactgcttggaaatttggaaat 1931  
Db 2161 ccaattataatcccaacataatactgaaataactgaaactgcttggaaatttggaaat 2220

QY 1932 ttctgatactccagtcgagccgagacacgctgag taaagga tgtgaggcagcagcagg 1991  
 |||||||  
 Db 2221 ttctgatactccagtcgagccgagacacgctgag taaagga tgtgaggcagcagcagg 2280  
 |||||||  
 QY 1992 aagacaacagaaacacacagagagcgctgtg cggcgctgagcctgtgctgaggttttgt 2051  
 |||||||  
 Db 2281 aagacaacagaaacacacagagagcgctgtg cggcgctgagcctgtgctgaggttttgt 2240  
 |||||||  
 QY 2052 gatggcactcgtgtgacccgctgagccgttcac gcaatagacgctgctgtgaggaaag 2111  
 |||||||  
 Db 2341 gatggcactcgtgtgacccgctgagccgttcac gcaatagacgctgctgtgaggaaag 2400  
 |||||||  
 QY 2112 gctgcccagccagctgtgtctcccgagacacgag cagatccacccctggcaccctcgt 2171  
 |||||||  
 Db 2401 gctgcccagccagctgtgtctcccgagacacgag cagatccacccctggcaccctcgt 2460  
 |||||||  
 QY 2172 gtttggctttttttccctgtgtgaaagaagaa cggcagcaccctctcaagctg 2231  
 |||||||  
 Db 2461 gtttggctttttttccctgtgtgaaagaagaa cggcagcaccctctcaagctg 2520  
 |||||||  
 QY 2232 ctccacacagacatctggagacaaaccttgagac gcatgcccagagagcctttgacg 2291  
 |||||||  
 Db 2521 ctccacacagacatctggagacaaaccttgagac gcatgcccagagagcctttgacg 2580  
 |||||||  
 QY 2292 gccccagagctaaagacacagagaaatcaaa tgcctcctacagcgtgacccaact 2351  
 |||||||  
 Db 2581 gccccagagctaaagacacagagaaatcaaa tgcctcctacagcgtgacccaact 2640  
 |||||||  
 QY 2352 ttctagtgagcagcgcccccacacactcctgag taccacacatcaacactgttct 2411  
 |||||||  
 Db 2641 ttctagtgagcagcgcccccacacactcctgag taccacacatcaacactgttct 2700  
 |||||||  
 QY 2412 ctcccaacagtgatctgatacttcttgaattat tttcttggatgataaacat 2471  
 |||||||  
 Db 2701 ctcccaacagtgatctgatacttcttgaattat tttcttggatgataaacat 2760  
 |||||||  
 QY 2472 ataaaaatttcattgagaaattctcaatgtat tctagttaa tagacagtttgaaac 2531  
 |||||||  
 Db 2761 ataaaaatttcattgagaaattctcaatgtat tctagttaa tagacagtttgaaac 2820  
 |||||||  
 QY 2532 ttgtctgagactgacttatacaataatctaac ccgacaaagatcatatccatgtgt 2591  
 |||||||  
 Db 2821 ttgtctgagactgacttatacaataatctaac ccgacaaagatcatatccatgtgt 2880  
 |||||||  
 QY 2592 gttagacattttatttacttgaactaacccagag acagtttcaatgtagcgaattgtgtc 2651  
 |||||||  
 Db 2881 gttagacattttatttacttgaactaacccagag acagtttcaatgtagcgaattgtgtc 2940  
 |||||||  
 QY 2652 cctctgttcaagctggaacagctcgtgactttca aaaaaccttgaataagctcccaag 2711  
 |||||||  
 Db 2941 cctctgttcaagctggaacagctcgtgactttca aaaaaccttgaataagctcccaag 3000  
 |||||||  
 QY 2712 tgtataaatggaacaattaggaatttaaac ttagatgatacatattgtgttcattt 2771  
 |||||||  
 Db 3001 tgtataaatggaacaattaggaatttaaac ttagatgatacatattgtgttcattt 3060  
 |||||||  
 QY 2772 ttctattttattttgtttaatgcaaacaggact taaatgaacttgaatcgtctgttt 2831  
 |||||||  
 Db 3061 ttctattttattttgtttaatgcaaacaggact taaatgaacttgaatcgtctgttt 3120  
 |||||||  
 QY 2832 agattatlaaaaaacttggtatctatacatatg ctcctgaggacttgcattca 2891  
 |||||||  
 Db 3121 agattatlaaaaaacttggtatctatacatatg ctcctgaggacttgcattca 3180  
 |||||||  
 QY 2892 cactacagagatgatactcactatgtagtcaat aacactgacagatgatttccagag 2951  
 |||||||  
 Db 3181 cactacagagatgatactcactatgtagtcaat aacactgacagatgatttccagag 3240  
 |||||||  
 QY 2952 ctgagatagtgtaatactatcactcattagagc tggaaagaatgacagtttctgtata 3011  
 |||||||  
 Db 3241 ctgagatagtgtaatactatcactcattagagc tggaaagaatgacagtttctgtata 3300  
 |||||||  
 QY 3012 cagctgtgtgtctttgtatgtgtgtactgtaca cagaagtggtgtcactgagcctg 3071  
 |||||||

Db 3301 cagctgtgtgtctttgtatgtgtgttactgtacac aagaagtggtgcactgagcctg 3360  
 |||||||  
 QY 3072 cgtgtgtgtccgtatgtaaaaaacctgtagc cctgaggttaagtaagcgttccattatg 3131  
 |||||||  
 Db 3361 cgtgtgtgtccgtatgtaaaaaacctgtagc cctgaggttaagtaagcgttccattatg 3420  
 |||||||  
 QY 3132 tttaagctggaattttctcccatgtaagttaaa acttaagtggttgcataat 3191  
 |||||||  
 Db 3421 tttaagctggaattttctcccatgtaagttaaa acttaagtggttgcataat 3480  
 |||||||  
 QY 3192 aaatggttaactataaaaaaaa 3217  
 |||||||  
 Db 3481 aaatggttaactataaaaaaaa 3506  
 |||||||  
 RESULT 6  
 293328  
 ID 293328 standard; cDNA: 2859 BP.  
 XX  
 AC 293328;  
 DT 04-JUL-2000 (first entry)  
 DE Mouse staufen cDNA.  
 XX  
 DE Staufen: RNA binding protein: RNA/DNA hybrid; virus: retrovirus;  
 KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;  
 KW acquired immune deficiency syndrome; protease; mouse; ss.  
 XX  
 OS Mus musculus.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 325..1788  
 FT /tag= a  
 FT /product= Staufen protein  
 PN CA2238656-A1.  
 XX  
 PD 22-NOV-1999.  
 XX  
 PF 22-MAY-1998; 98CA-2238656.  
 XX  
 PR 22-MAY-1998; 98CA-2238656.  
 XX  
 PA (UYMO-) UNIV MONTREAL.  
 XX  
 PI Duchaine T, Wickham L, Cohen EA, Luo M, Desrosiers L;  
 PI Mouloud AJ;  
 DR WPI: 2000-246924/22.  
 DR P-PSDB: Y83024.  
 XX  
 PT Isolated mammalian or Caenorhabditis elegans staufen proteins useful  
 PT for treating retroviral infections especially human immunodeficiency  
 PT virus infections (HIV)  
 XX  
 PS Claim 4; Figure 1c; 96pp; English.  
 XX  
 CC Staufen is a RNA binding protein which interacts with double stranded  
 CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA  
 CC transport and localization. Mammalian SP contain multiple ds  
 CC RNA-binding domains. Chimeric proteins comprising portions of  
 CC staufen with proteins with RNase or protease activity can be used to  
 CC treat virus infections. The RNase or protease activity of the fusion  
 CC protein prevents proper maturation of the virus. Mammalian staufen  
 CC proteins seem to recognise double stranded RNA structure rather than  
 CC any sequence specific position. 2 bacterially-expressed fusion  
 CC proteins used in an RNA-binding assay (his/hsp (human SP) and  
 CC MBP/hsp (murine SP) both proteins strongly bound double stranded  
 CC RNA. Both fusion proteins also directly bound labeled ds RNAs and  
 CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the  
 CC treatment of RNA virus infections, especially those caused by



AC X39991;  
 XX 02-JUL-1999 (first entry)  
 DT XX  
 DE Prostate cancer associated gene.  
 XX  
 XX Cancer associated antigen; diagnosis; research; treatment; human;  
 KM breast cancer; colon cancer; gastric cancer; lung cancer;  
 KM prostate cancer; ss.  
 XX  
 OS Homo sapiens.  
 PN W09904265-A2.  
 PD 28-JAN-1999.  
 XX  
 PF 15-JUL-1998; 98MO-US14679.  
 XX  
 PR 22-JUN-1998; 98US-0102322.  
 PR 17-JUL-1997; 97US-0896164.  
 PR 10-OCT-1997; 97US-0061599.  
 PR 10-OCT-1997; 97US-0061765.  
 PR 10-OCT-1997; 97US-0948705.  
 PR 11-OCT-1997; 97GB-0021697.  
 XX  
 XX (LUDM-) LUDMIG INST CANCER RES.  
 PA  
 PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;  
 PI Pfundschuh M, Sahin U, Scanlan MJ, Stockert E;  
 PI Tureci O;  
 DR WPI; 1999-132448/11.  
 XX  
 XX New isolated cancer associated nucleic acids and polypeptides -  
 PT isolated using sera from cancer patients, used to develop products  
 PT for the diagnosis, monitoring or treatment of cancers  
 PS  
 PS Claim 67: Page 625-626; 787pp: English.  
 XX  
 CC The invention relates to a method for diagnosing a disorder characterised  
 CC by expression of a human cancer associated antigen precursor coded for by  
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
 CC biological sample isolated from a subject with an agent that specifically  
 CC binds to the NAM, an expression product or a fragment of an expression  
 CC product complexed with an HLA molecule; and (b) determining the  
 CC interaction between the agent and the NAM or the expression product as a  
 CC determination of the disorder. The products and methods can be used in  
 CC the diagnosis, monitoring, research, or treatment of conditions  
 CC characterised by the expression of various cancer associated antigens.  
 CC The invention provides nucleic acid sequences and encoded polypeptides  
 CC which are cancer associated antigen precursors, expressed in human breast  
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
 CC lung cancer.  
 CC  
 XX  
 XX Sequence 1127 BP; 349 A; 265 C; 307 G; 203 T; 3 other;  
 SQ  
 Query Match 34.4%; Score 1107.2; DB 20; Length 1127;  
 Best Local Similarity 99.4%; Pred. No. 1.3e-272;  
 Matches 1120; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 700 ttaaacggaacttgcgttgaatttcgaggtgcccggagagtgccaccacatga 759  
 DB ttaacggaacttgcgttgaatttcgaggtgcccggagagtgccaccacatga 240  
 QY 760 agaacttctgacaaagtttcggttgggaggttttgggggaggttgaaggaaaga 819  
 DB agaacttctgacaaagtttcggttgggaggttttgggggaggttgaaggaaaga 300  
 QY 820 agagaattcacaagaataatgcgcacatagcttcttgaaggagctgaagaattacccg 879  
 DB agagaattcacaagaataatgcgcacatagcttcttgaaggagctgaagaattacccg 360  
 QY 880 cccctctcagtttgaacagtaagcctagaatcaaaaaaagaacacacatagta 939  
 DB cccctctcagtttgaacagtaagcctagaatcaaaaaaagaacacacatagta 420  
 QY 940 agccacagacaaagccagatatgcccagggaggttcacatccgatccgactgcccaga 999  
 DB agccacagacaaagccagatatgcccagggaggttcacatccgatccgactgcccaga 480  
 QY 1000 tccagcagcacaataaaggaagagccagagtaacacgctcctcacagaagcagagctcc 1059  
 DB tccagcagcacaataaaggaagagccagagtaacacgctcctcacagaagcagagctcc 540  
 QY 1060 cgcgcgcagggaggttctgtatgacgttgaaggcttggaaacacactgcagaagagcgg 1119  
 DB cgcgcgcagggaggttctgtatgacgttgaaggcttggaaacacactgcagaagagcgg 600  
 QY 1120 gacccaagaagaagtgcccaagcgaatgcacgcagagaacacatgcttggatcttgggt 1179  
 DB gacccaagaagaagtgcccaagcgaatgcacgcagagaacacatgcttggatcttgggt 660  
 QY 1180 tcaaatcccgagcgagccagccacacacacacacacacacacacacacacacacacac 1239  
 DB tcaaatcccgagcgagccagccacacacacacacacacacacacacacacacacacac 720  
 QY 1240 taaagaacacagggagttggaagaagaatccttttgaacctgcttgggagtaaa 1299  
 DB taaagaacacagggagttggaagaagaatccttttgaacctgcttgggagtaaa 780  
 QY 1300 atggagtagtaataaagaagatgatttcagatgcttatcaatgtaacacagctgc 1359  
 DB atggagtagtaataaagaagatgatttcagatgcttatcaatgtaacacagctgc 840  
 QY 1360 ctctggaattcttccatgctgcccaggttgcggaggttgaaggatgaacagac 1419  
 DB ctctggaattcttccatgctgcccaggttgcggaggttgaaggatgaacagac 900  
 QY 1420 atcacacaaagatttaccagagcagcttcgaatccttgcacaaagccagtaactgca 1479  
 DB atcacacaaagatttaccagagcagcttcgaatccttgcacaaagccagtaactgca 960  
 QY 1480 tgaagccgagaggttctgtatgaggggacacccgcacaaagccagacacattttaaag 1538  
 DB tgaagccgagaggttctgtatgaggggacacccgcacaaagccagacacattttaaag 1020  
 QY 1539 aataaactcttcaagccagccagccacacacacacacacacacacacacacacacacac 1598  
 DB aataaactcttcaagccagccagccacacacacacacacacacacacacacacacacac 1080  
 QY 1599 gactatcttccagagtcagagatccagatttgaattacaagaactt 1645  
 DB gactatcttccagagtcagagatccagatttgaattacaagaactt 1127  
 RESULT 8  
 X40002  
 ID X40002 standard; DNA; 946 BP.  
 XX  
 AC X40002;  
 XX  
 DT 02-JUL-1999 (first entry)  
 XX



PR 10-OCT-1997; 97US-0061599.  
 PR 10-OCT-1997; 97US-0061765.  
 PR 10-OCT-1997; 97US-0948705.  
 PR 11-OCT-1997; 97GB-0021697.  
 XX  
 PA (LUDW-) LUDWIG INST CANCER RES.  
 XX  
 PI Chen Y, Gout I, Gure A, O'Hare M, Ohta Y, Old LJ;  
 PI Pfunderschuh M, Sahin U, Scanlan MJ, Stockert E;  
 PI Tureci O;  
 XX  
 DR WPI; 1999-132448/11.  
 XX  
 PR New isolated cancer associated nucleic acids and polypeptides -  
 PR isolated using sera from cancer patients, used to develop products  
 PR for the diagnosis, monitoring or treatment of cancers  
 XX  
 PS Claim 67; Page 629; 787pp; English.  
 XX  
 CC The invention relates to a method for diagnosing a disorder characterised  
 CC by expression of a human cancer associated antigen precursor coded for by  
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
 CC biological sample isolated from a subject with an agent that specifically  
 CC binds to the NAM, an expression product or a fragment of an expression  
 CC product complexed with an HLA molecule; and (b) determining the  
 CC interaction between the agent and the NAM or the expression product as a  
 CC determination of the disorder. The products and methods can be used in  
 CC the diagnosis, monitoring, research, or treatment of conditions  
 CC characterised by the expression of various cancer associated antigens.  
 CC The invention provides nucleic acid sequences and encoded polypeptides  
 CC which are cancer associated antigen precursors expressed in human breast  
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
 CC lung cancer.  
 XX  
 SQ Sequence 844 BP; 251 A; 196 C; 234 G; 156 T; 7 other;

Query Match 23.5%; Score 755.6; DB 20; Length 844;  
 Best Local Similarity 96.3%; Pred. No. 5.3e-183;  
 Matches 812; Conservative 0; Mismatches 26; Indels 5; Gaps 4;

QY 682 aagtggttgagatgacttaaaaggaacttgctgtgaattcgaagtgccgggaga 741  
 |||||||  
 DB 1 aagtggttgagatgacttaaaaggaacttgctgtgaattcgaagtgccgggaga 60  
 QY 742 gtggccacccacatgaagaactttgtgacaaagtttcgtgtgggaattgtgggg 801  
 |||||||  
 DB 61 gtggccacccacatgaagaactttgtgacaaagtttcgtgtgggaattgtgggg 120  
 QY 802 aagtgtaagggaaagcaagaatttcaagaagaaatgcgcacatagctgttcgag 861  
 |||||||  
 DB 121 aagtgtaagggaaagcaagaatttcaagaagaaatgcgcacatagctgttcgag 180  
 QY 862 agctgaagaagtataccgcccctgctgctgagtgtaacgagtaaaagcctagaatcaaaaga 921  
 |||||||  
 DB 181 agctgaagaagtataccgcccctgctgctgagtgtaacgagtaaaagcctagaatcaaaaga 240  
 QY 922 aaacaaaccatagtcacagccacagaacagccagaataatggtccagggatcaatccga 981  
 |||||||  
 DB 241 aaacaaaccatagtcacagccacagaacagccagaataatggtccagggatcaatccga 300  
 QY 982 tttagcgcactggtccagatccagcaagcaaaaaggaaggaagcaggaatcacagctcc 1041  
 |||||||  
 DB 301 tttagcgcactggtccagatccagcaagcaaaaaggaaggaagcaggaatcacagctcc 360  
 QY 1042 tcacagagcagagcctcccgccgagcagggaggtttgtatgcagtgtaaggttgaacc 1101  
 |||||||  
 DB 361 tcacagagcagagcctcccgccgagcagggaggtttgtatgcagtgtaaggttgaacc 420  
 QY 1102 acactgcagagaaggaagggacacaaagaaggttggtccaaagcgcgaagcgcgagaca 1161  
 |||||||  
 DB 421 acactgcagagaaggaagggacacaaagaaggttggtccaaagcgcgaagcgcgagaca 480

QY 1162 tgcctgagatccttggtttccaaggtccgcagcgccagccaccacaaaccgcactaagt 1221  
 |||||||  
 DB 481 tgcctgagatccttggtttccaaggtccgcagcgccagccaccacaaaccgcactaagt 540  
 QY 1222 cagagagaagaagaccacataaagaacccgggagtgagaagaagaatgaaccttttgaac 1281  
 |||||||  
 DB 541 cagagagaagaagaccacataaagaacccgggagtgagaagaagaatgaaccttttgaac 600  
 QY 1282 ctgcctctggggatgaaatggagcagtaataaagaagatgagttcagatgcttatc 1341  
 |||||||  
 DB 601 ctgcctctggggatgaaatggagcagtaataaagaagatgagttcagatgcttatc 660  
 QY 1342 taagtcataagcagctgctgctggaattcctccatggtgcccagagtcgcccagctg 1401  
 |||||||  
 DB 661 taagtcataagcagctgctgctggaattcctccatggtgcccagagtcgcccagctg 720  
 QY 1402 taagtgtaaggaagacatcaccaagaattttacccgggagctccgaatctgccc 1460  
 |||||||  
 DB 721 taaggaagtgtaaggaagacatcaccaagaattttacccgggagctccgaatctgccc 779  
 QY 1461 aagggcagcgttaactgcacatgatacccgagagtggtgatgaggagcactgcgcaca 1520  
 |||||  
 DB 780 angcngng--tactgcacatgatagcc--anagttgtgatgaggagcactgtgccca 836  
 QY 1521 gcc 1523  
 |||  
 DB 837 ggc 839

RESULT 10  
 X39992  
 ID X39992 standard; DNA; 773 BP.  
 XX  
 AC X39992;  
 XX  
 DT 02-JUL-1999 (first entry)  
 XX  
 DE Prostate cancer associated gene.  
 XX  
 KW Cancer associated antigen; diagnosis; research; treatment; human;  
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
 KW prostate cancer; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO9904265-A2.  
 PD 28-JAN-1999.  
 XX  
 PF 15-JUL-1998; 98WO-US14679.  
 XX  
 PR 22-JUN-1998; 98US-0102322.  
 PR 17-JUL-1997; 97US-0896164.  
 PR 10-OCT-1997; 97US-0061599.  
 PR 10-OCT-1997; 97US-0061765.  
 PR 10-OCT-1997; 97US-0948705.  
 PR 11-OCT-1997; 97GB-0021697.  
 XX  
 PA (LUDW-) LUDWIG INST CANCER RES.  
 XX  
 PI Chen Y, Gout I, Gure A, O'Hare M, Ohta Y, Old LJ;  
 PI Pfunderschuh M, Sahin U, Scanlan MJ, Stockert E;  
 PI Tureci O;  
 XX  
 DR WPI; 1999-132448/11.  
 XX  
 PR New isolated cancer associated nucleic acids and polypeptides -  
 PR isolated using sera from cancer patients, used to develop products  
 PR for the diagnosis, monitoring or treatment of cancers  
 XX  
 PS Claim 67; Page 626; 787pp; English.  
 XX  
 CC The invention relates to a method for diagnosing a disorder characterised





```
QY 2719 attggaacattgaggaatttaaacttagatgacatttggttcatttttcaatt 2778
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 671 ATGAGCAATTTGGAATTTAACTTGAATGATCATTTGGTCAATTTATTTTCAT 612
QY 2779 ttatttttgtaaagcaaacagac-ttaaatgaacttgatctgttttaagatta 2837
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 611 TTTATTTTGTATTAATCAACAGACTTTAAATGAACCTTTGATCTCTGTTTAAAGATTA 552
QY 2838 ttaaaaacatttgatcatcatatcatatgctttgaggacttagctttactaacctac 2897
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 551 TTAATAAACATTTGTATCTATACATATGAGCTTTGAGGACTTACCTTACATACATAC 492
QY 2898 agagatgacatccatgtagttagtataaacctgcagagatcttccaaagtcgcgat 2957
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 491 AGGATATGATCTTCATGTATCATATTAACCTGCAGAGATTTTCCAGAGTCTCGAT 432
QY 2958 actgttaattacatctccattagagctgaaagaatgacctagcttctglatacagctg 3017
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 431 ACGTTAATTAATACATCTCCATTTAGGCGTGAAAGAAATGACCTACGTTTCTGTATACAGCTG 372
QY 3018 tgtgtcttttgatggtgtgttactgtacacagaagtggtgacactgagctcgcgtgtg 3077
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 371 TGTGTGTTTGTATGTTGTGTACTGTACACAGAGTGTGTGCACTGAGGCTTGCCTGTG 312
QY 3078 gtccgataggaacacctggtgagcccgcgagtgtaagtaactgcttccattatgattag 3137
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 311 GTCCGATATGAAAGACCTGTGAGCCCTGCGAGTTAAGTACTGCTTCCATTTATGTTTACG 252
QY 3138 ctggaatttctcccccacatggaatgtaagtaaaacttaagtggttgcatacaataatg 3197
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 251 CTGGAATTTTTCCTCCCATGGAATGTAACTTAAGTGTGTTGTCATCAATTAATG 192
QY 3198 taactactaaa 3207
      ||||| |||||
Db 191 TAATACTAAA 182

RESULT 12
X39993/c
ID X39993 standard; DNA: 794 BP.
XX
AC X39993;
XX
DT 02-JUL-1999 (first entry)
XX
DE Prostate cancer associated gene.
XX
KM Cancer associated antigen; diagnosis; research; treatment; human;
KM breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KM prostate cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ,
PI Pfrendschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
```

```
DR WPI; 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Claim 67; page 626; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
SQ Sequence 794 BP: 273 A; 141 C; 152 G; 206 T; 22 other;
```

```
Query Match 15.6%; Score 501.4; DB 20; Length 794;
Best Local Similarity 96.3%; Pred. No. 3.5e-118;
Matches 517; Conservative 0; Mismatches 18; Indels 2; Gaps 1;

QY 2673 tccctgagcttcaaaaacactgtaaaagtcctcc--acagttgataaatggaacatt 2730
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 718 TCCGAGACTTTCAAAACCCCTGAATTAAGTCTCCGCCACAGTTGNATTAATGGCCNATTT 659
QY 2731 aggaattttaaacttgaatgacatttggttccattttatcatttattttgtt 2730
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 658 AGGAATTTTAAACMTTATGATGTTGTTCCATTTTATTTTATTTATTTTGGT 599
QY 2791 aatgcaaacaggacttaaaacttgatctcgtttttaaagaatttaaacaacttg 2850
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 598 AATGCANACAGGNCCTTAAGTAAGACTTTGATCTCTGTTTAAAGTTTAAAAACATGG 539
QY 2851 tgaataacataatggtcctctgaggaacttgacttccactaacactaacagatatgactc 2910
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 538 TGNATCTATACATATGCTCTGTGAGACTTACCTTTCACATACAGATATGATCTC 479
QY 2911 catgtacatataaacctgcagagtgatctccagagtgctgcatactgtaataca 2970
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 478 CATGTAGTCCATATATAACCTGCAGAGTGATTTTCCANAGTCTCGATATCTGTTAATTACA 419
QY 2971 tctccattagggctgaaaagatgactagttctctgtatacagctgtgtgctttgat 3030
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 418 TCTCCNTTAAAGGCTGAAGAAAGATGACCTGCTTCTGTATACANCTGTGCTTTTGAT 359
QY 3031 gttgtgtaactgacacagaagtgtgtcacatgaggtcgtggtgtgtccgtatggaaa 3090
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 358 GTTGTGNTACTGNNACACAGAGTGTGNCNTGAGGCTCTCGTGTGCGTATGAGAA 299
QY 3091 acctgtgagccctgcaggttaagtactgcttccattcaattgttaagctgtgaatttctc 3150
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 298 GCTGTGTAGCCCTCGAGTTAAGTACCTCTTCATCTCAATTTTAAAGCTGGAATTTTCT 239
QY 3151 ccccatggaatgtaagtaaaacttaagtggttgcatacaataatgtaatacaaa 3207
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 238 CCCCATGCAATGTAAAGTAAACTTAAGTGTGTGTCATCAATNAATGGTAATACATAA 182

RESULT 13
279981/c
ID 279981 standard; cDNA: 660 BP.
XX
AC 279981;
```



```

|||||
Db 1 ggcatttaccattccattccattccatttacttacttacttcaatgagaaacttctgtgag 60
QY 523 gaagcaatttaattgcaagaagaagaagaagaagctgtgcaaacagatgctgtcca 582
Db 61 gacagcaatttaattgcaagaagaagaagaagaagctgtgcaaacagatgctgtcca 120
QY 583 aagcgttgaagatcctgcaaatgagccctgtccagagagctggaagtgatgaagaag 642
Db 121 aagcgttgaagatcctgcaaatgagccctgtccagagagctggaagtgatgaagaag 180
QY 643 aatccgaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 702
Db 181 aatccgaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 240
QY 703 aagcgaactgtgctgtgaatttgcgagctgcccggagagatgccccacatgaaga 762
Db 241 aagcgaactgtgctgtgaatttgcgagctgcccggagagatgccccacatgaaga 300
QY 763 acttctgaccgaagttctggtgagctgtgtgaggaagtgaaaggaagaagaaga 822
Db 301 acttctgaccgaagttctggtgagctgtgtgaggaagtgaaaggaagaagaagaaga 360
QY 823 agatttcaaaagaaaatg-cgcgcatagtctgtcttgaaggagctgaagaagttaccgcc 881
Db 361 agatttcaaaagaaaatg-cgcgcatagtctgtcttgaaggagctgaagaagttaccgcc 420
QY 882 ctgcctgcaatgaac---gagtaaaagcctagaatcaaaaagaaga 923
Db 421 ctgcctgcaatgaacccgaagttaaagcctanaatccaatgaaga 465

RESULT 15
280703/C
ID 280703 standard: cDNA, 708 BP.
XX
AC 280703;
XX
DT 07-APR-2000 (first entry)
XX
DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:787.
XX
KW Human; gene expression product; diagnosis; tumour; colon cancer;
KM colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
XX hyperplasia; ds.
XX
OS Homo sapiens.
XX
PN M09964576-AZ.
XX
PD 16-DEC-1999.
XX
PF 09-JUN-1999; 99MO-IB01062.
XX
PR 10-JUN-1998; 98US-0088801.
XX
PA (FARB ) BAYER CORP.
XX
PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catino TJ, Dertl A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;
XX
DR WPI; 2000-087220/07.
XX
PT Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer
XX
PS Claim 15; Page 444; 469pp; English.
XX
CC 279917 to 280766 represent double stranded cDNA clones isolated from the
CC human colorectal adenocarcinoma (colon cancer) cell line SW480. The

```

```

CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g. carcinomas
CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used
CC for developing agents for the diagnosis and treatment of disorders
CC involving unwanted cell proliferation, such as neoplasia, dysplasia or
CC hyperplasia.
XX
SQ Sequence 708 BP; 226 A; 127 C; 143 G; 191 T; 21 other;

Query Match          10.2%; Score 327.8; DB 21; Length 708;
Best Local Similarity 82.8%; Pred. No. 6.2e-74;
Matches 515; Conservative 0; Mismatches 83; Indels 24; Gaps 12;

QY 2446 tttcttcttgaatgatgaac--tatataaatcttcatatgaagattctc---aat 2500
Db 622 TTTCNTTNGATGAAATGGCCCTAATAAATAATTCNTTGAATAATTTTCCCAATG 563
QY 2501 tgaatcgaatgaatgaacagatgtggaact--gtctgagatgaacttacc--aat 2555
Db 562 GGATCCAAAGTTAAATACCCAGTTGGAACCTTGTTAAAGCTGCTTTAAACCAATA 503
QY 2556 aatcacaagac--aaagatcatacattgtatgtgttgaacatttatttcatgt 2613
Db 502 ATTGNAACCGCGCAAAAGTTCAAAACCNAGGGAAGNGNATTAACNNATTTAATTCATG 443
QY 2614 actaaccc--aggaagcttcaatgaatgaatgtgtccctt--gttcaagctgaaga- 2669
Db 442 NCTTACCCCGAGGCGCACTTCAGTAACCAAAATGGGGGCTTCTGCGGTTCAGCTGAAC 383
QY 2670 cagtcctggaacttcaaaaacctgaa-taagtctcca-cagttgataaattgagcaa 2727
Db 382 CAGTCTGGAAGTTCCAAACCCGTAATTAAGTCCACAGTGTGTAATATGGCA 323
QY 2728 tttagaatlttaaaccttgaatgaac-----tttgttccatttatttcaattt 2782
Db 322 TTTAGGAATTTTAAACCTTTTAAATGATGATGATGATGATGATGATGATGATGAT 263
QY 2783 ttttgttaatgaagaagaagacttaaatgaacttgaatctgttcttaagaattataa 2842
Db 262 TTTTGTAAATGCAAAACAGGACTTAAATGAACCTTGATCTCTGTTTAAAGATTATTA 203
QY 2843 aaacattgtatcatatcatatgctcttgaggaactgaattcacttaacatacagat 2902
Db 202 AAACATTGTGATCTATCATATGCTCTTGAGGACTTACACTACACTACAGGAT 143
QY 2903 atgattccatgtagtccataataaactgcagagtgatttccagagtgctcgaactgt 2962
Db 142 ATGATCTCCAGTGTAGTCCATTAACCTGCAGAGTGAATTTCCAGAGTCTCGATACTGT 83
QY 2963 taattacatcctcattagggctgaagaagaatgacctacgcttctgatacag-ctgtgtt 3021
Db 82 TAATTACATCTCCATTTAGGCTGAAGAAAGATGACCACGTTTGTATACAGCTGTGT 23
QY 3022 gctttagatgtgtgtactgt 3043
Db 22 GCTTTGATGTGTGTACTGT 1

```

Search completed: April 5, 2001, 08:30:05  
Job time: 59249 sec

• • •



alignment\_block:  
US-09-316-048-11 x 293328 ..

Align seg 1/1 to: 293328 from: 1 to: 2859

```

1 MetTyrLysProValAspProHisSerArgMetGlnSerThrTyrSerTy 17
  |||||
325 ATGTATAAAGCCCGTGGACCCCTCACTCGGATGCAGTCCACCTACAGCTA 374
  |||||
17 rGlyMetArgGlyGlyAlaTyrProProArgTyrPheTyrProPhePro 34
  |||||
375 TGGCATGCGTGGAGTGCCTATCCGCCAGATCTTTACCCATTTCACAG 424
  |||||
34 aLProProLeuLeuTyrGlnValGlnLeuSerValGlyGlnGlnInPhe 50
  |||||
425 TCCCACTTTACTCTACCAAGTTGAGCTCTCCGTGGGCGGACAGCAAGT 474
  |||||
51 AsnGlyLysGlyLysMetArgProProValLysHisAspAlaProAla 67
  |||||
475 AATGGGAAGGAAGAAAGATGAGACCCCGTGAACAGATGCCCTGCCCG 524
  |||||
67 gAlaLeuArgTThrLeuGlnSerGluProLeuProGluArgLeuGluVal 84
  |||||
525 TCGCGTGAAGGACTCTGACAGTGAACCCCTGCCAGAAAGCTTGAGGTA 574
  |||||
84 sngLysArgGluAlaGluGluGlnLysAsnLeuAsnLysSerGluLeuSer 100
  |||||
575 ATGGAAAGAGAGAGAGAGAAACCTCAATTAATCGGAATTAACCCAA 624
  |||||
101 ValPheGluLysLeuLysAsnLysArgAsnLeuProValAsnPheGluVal 117
  |||||
625 GTGTTGAATTCGCTGAAGCGGAATTTGCCGTGATTTTGAAGTGGCG 674
  |||||
117 aArgGluSerGlyProProHisMetLysAsnPheValThrArgValSer 134
  |||||
675 CCGGAGAGTGGCCACACACATGAAGAACTTGTACACAGGTTTCAG 724
  |||||
134 aLgLyLysPheValGlyGluGlyGlnLysSerLysLysLysLysLys 150
  |||||
725 TTGGGAATTTGTAGGGAGAGAGAGAGAAAGCAAGAGATCTCCAG 774
  |||||
151 LysAsnAlaAlaArgAlaValLeuGlnGlnLeuArgArgLeuProPro 167
  |||||
775 AAGAATAGCGGCCGAGGCTGTTCGAGCAGCTTAGAGAGCTGCACCCCT 824
  |||||
167 uProAlaValGluArgValLysProArgLysLysLysSerGluProT 184
  |||||
825 CCTGCTGTGACGAGTGAAGCCCAAGATCAAGAAAGAAAGTCAACCCA 874
  |||||
184 hrcYsLysThrAlaProAspTyrGlyGlnGlyMetAsnProLysSerArg 200
  |||||
875 CCTGCAGACAGCCCGGATTTATGCCAAGGATGATCTTATTAGTAGA 924
  |||||
201 LeuAlaGlnIleGlnGlnAlaLysLysGlyLysGluProGluTyrMetLe 217
  |||||
925 CTTCGACAGATCCAGCAGGCAAGGAGGTGATCCAAATTAAGAAAGTGGCC 1074
  |||||
217 uLeuThrGluArgGlyLeuProArgArgArgGluPheValMetGlnVal 234
  |||||
975 CCTTACAGAAACGAGTCTTCCAGCTGCAGGAGTTTGTGATGCAGGTAA 1024
  |||||
234 yValGlyHisHisThrAlaGluGlyValGlyThrAsnLysLysValAla 250
  |||||
1025 AGGTGTGGCATCACTGCAAGAGAGGTGGTATCCAAATTAAGAAAGTGGCC 1074
  |||||
251 LysArgAsnAlaAlaGluAsnMetLeuGluLysLysLysLysValPyr 267
  |||||
1075 AAGCGTAATGCTGCTGAGAAACATGCTGAGATCTCGGGGTCAAAATTCC 1124
  |||||
267 oGlnAlaGlnProAlaLysProAlaLeuLysSerGluGluLysThrPro 284
  |||||
1125 CCAGCGCAGCGCTGCCAAGCAGCACTCAAAATCAGAAAGAAAGACTCCAG 1174

```

```

284 aLysLysProGlyAspGlyArgLysValThrPhePheGluProSerPro 300
  |||||
1175 TAAAGAAACCGAGACGACGAGAGAAAGTAAAGCTTTTGTGACCTACCCCT 1224
  |||||
301 GlyAspGluAsnGlyThrSerAsnLysAspGluGluPheArgMetProTy 317
  |||||
1225 GGGATGAAATGAGACTAGTAAACAAGACGAGAGACTTACAGATGCTTAA 1274
  |||||
317 rLeuSerHisGlnGlnLeuProAlaGlyLysLeuProMetValProGlu 334
  |||||
1275 TCTTACCCATCAGCAGCTGCGCAGTGAATTCCTCCATGCTGCCGAAG 1324
  |||||
334 aAlaGlnAlaValGlyValLysGlnGlyHisHisThrLysAspPheThr 350
  |||||
1325 TTGCCAGGCTGCTGGGGTGTAGTCAAGACACCCACCAAGATTTCACC 1374
  |||||
351 ArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetIleAla 367
  |||||
1375 AGGCAAGCTCCAAATCTCGCAAGGCAAGGTAAGTCCATGATAGCCCG 1424
  |||||
367 gGluLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLys 384
  |||||
1425 AGAGTTGTGTACGGGGGCACTCCGCCACAGCCGAGACCATTTTAAAGA 1474
  |||||
384 eArgAsnLysSerSerGlyHisValProHisGlyProArgThrArgPro 400
  |||||
1475 GTTACATCTCTTACAGGCCAGCTACCCATGAGCTGCGCATAGACCTCT 1524
  |||||
401 GluGlnLeuTyrThrLeuSerArgAlaGlnGlyPheGlnValGluTyrLy 417
  |||||
1525 GAGCAACTGTACTACTTTCAGAGCCCAAGGATTCAGAGTTGAATACAA 1574
  |||||
417 sAspPheProLysAsnAsnLysAsnLysValSerLeuIleAsnCys 434
  |||||
1575 AGATTTTCCCAAGAACACAGAGAGTGTATCTTCATCACTACCTGCT 1624
  |||||
434 eSerGlnProProLeuValLysHisGlyIleGlyLysAspValGluSer 450
  |||||
1625 CCTCACAGCCGCTCTGTCGTCAGTCATGCGCATGCGCAAGATGGAAGTCC 1674
  |||||
451 CysHisAspMetAlaAlaLeuAsnIleLeuLysLeuLeuSerGluLeu 467
  |||||
1675 TTTCATGATATATGCTGCTGACGATGATTTAAAGCTGCTGAGTTGGA 1724
  |||||
467 pGlnGlnSerThrGluMetProArgThrGlyAsnGlyProValSerAla 484
  |||||
1725 CCNACAGACGACAGAGATGCGCAAGAACAGAAATGAGCAGTTTCAGCGT 1774
  |||||
484 ysgLYArGcys 487
  |||||
1775 GCGGAGAGTGC 1785

```

seq\_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1999.DAT:X90786

seq\_documentation\_block:  
ID X90786 standard; DNA: 3190 BP.

XX X90786;

AC 13-JAN-2000 (first entry)

XX XX

DE DNA encoding human staufer protein.

XX Human staufer protein; hStau; protein lysate; lung; kidney; testis;  
ovary; brain; heart; telomerase RNA; complex; catalytic subunit;  
telomerase assembly; maturation; transport; regulation; interaction;  
inhibitory agent; antibody; immunoassay; ds.

OS Homo sapiens.

XX XX

Key Location/Qualifiers

PH 345..1835

FT CDS

```
FT      /tag= a
FT      /product= "Human staufen protein"
FT      /note= "Binds to telomerase RNA"
FT      polyA_signal 3172..3177
FT      /*tag= b
PN      W09951255-A1.
PD      14-OCT-1999.
XX
XX      06-APR-1999; 99WO-US07533.
PF      06-APR-1998; 98US-0080783.
PR      (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
PA      Greider CW, Le S;
XX      WPI; 1999-620168/53.
DR      P-PSDB; Y26335.
XX
XX      Human staufen polypeptide useful in methods for identifying telomerase
PT      inhibitors
XX
PS      Claim 13; Page 41-42; 50pp; English.
XX
XX      The present sequence is an isolated DNA encoding human staufen (hStau)
CC      protein. The hStau protein was detected in protein lysates from lung,
CC      kidney, testis and ovary, but not in brain and heart. It binds
CC      specifically to human telomerase RNA and can complex with the
CC      telomerase catalytic subunit. The hStau protein may play a role in
CC      telomerase assembly, maturation, transport and regulation. Interaction
CC      between hStau and telomerase can be inhibited by administering inhibitory
CC      agents to the cell. These telomerase inhibitors can be identified using
CC      hStau protein. Antibodies generated against hStau can be used in various
CC      immunassays.
SQ      Sequence 3190 BP; 880 A; 788 C; 736 G; 786 T; 0 other;

alignment_scores:
      Quality: 2333.00      Length: 489
      Ratio: 4.922      Gaps: 1
Percent Similarity: 96.933      Percent Identity: 91.002

alignment_block:
US-09-316-048-11 x X90786 ..

Align seg 1/1 to: X90786 from: 1 to: 3190

1 MetTyrIysProValAspProHisSerArgMetGlnSerThrTyrSerty 17
|||||
366 ATGTATAGCTGTGACCTTACTCTCGATGGGCTCCACCTTAACCTA 415
17 rGlyMetArgIyGlyAlaTyrProProArgTyrPheTyrProPhePro 34
|:::
416 CAACATGAGAGAGAGTCTTATCCCGAGGTACTTTTACCATTTCAG 465
34 aLProProLeuLeuTyrGlnValGlnLeuSerValGlyGlnGlnPhe 50
|||||
466 TTCACCTTACTTATCAAGTGAACCTTCTGTGGAGGAGACAGCAATTT 515
51 AsnGlyLysGlyLysMetArgProProValLysHisAspAlaProAla 67
|||||
516 AATGGCAAGAAAGACAAAGACAGAGCTGCGAAACGATGCTGCGCAA 565
67 gAlaLeuArgThrLeuGlnSerGluProLeuProGluArgLeuGluVal 84
|||||
566 AGCGTTGAGGATCTCTGCAATGAGACCTCTCCACAGAGGCTGGAGGTGA 615
84 snGlyArgGluAlaGlnGluLysLeuAsnLysSerGluIleSerGln 100
|||||
616 ATGGAAGAGATCCGAGACAGAAATATCTCAATTAATCTGAATAAGTCAA 665
```

```
101 ValPheGluIleAlaLeuLysArgAsnLeuProValAsnPheGluValAl 117
|||||
666 GTGTTTGAGATTGCACCTTAACGGAATCTTGGTGAATTCGAGAGTGGC 715
117 aArgGluSerGlyProProHisMetLysAsnPheValThrArgValSer 134
|||||
716 CCGGAGAGAGTGGCCACCACATGAAAGACTTGTGTGACCAAGTTTGGG 765
134 aLgLyGluPheValGlyGlnGluGlyLysSerLysIleSerLys 150
|||||
766 TTGGGAGATTGTGGGGAGGTGAAGGAAAGCAAGAGATTTTCAAG 815
151 LysAsnAlaAlaArgAlaValLeuGlnLeuArgArgLeuProProLe 167
|||||
816 AAAATGCGCGCATAGCTCTTCTTGTAGAGAGTGAAGAGTTTACGCCCT 865
167 uProAlaValGluArgValLysProArgIleLysLysLysSerGlnPro 184
|||||
866 GCTGACAGTTGACGAGTAAAGCCTAGATCAAAAAGAAACCAACCA 915
184 hrcysLys.....ThrAlaProAspTyrGlyGlnGlyMetAsnProIle 198
|||||
916 TAGTCAGCCACAGACAGCCAGATATGGCCAGGAGATCTCGATT 965
199 SerArgLeuAlaGlnIleGlnGlnAlaLysLysGlyLysGluProGlu 215
|||||
966 AGCGCACATGGCCACATGATCCAGAGCAAAAAGGAAAGAGCCAGAGTA 1015
215 rMetLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMet 232
|:::
1016 CACGCTCCTCACAGAGCGAGGCTCCCGCGGAGGAGTGTGTGATGC 1065
232 lValAlaValGlyLysHisThrAlaGlnGluGlyValGlyThrAsnLys 248
|||||
1066 AGGTGAAGTTGGAACCAACACCTGCAGAAAGAACGGGACCAACAGAG 1115
249 ValAlaLysArgAsnAlaAlaGlnLysMetLeuGlnIleLeuGlyPhe 265
|||||
1116 GTGGCCAAAGCCGATGCAAGCCAGACATGCTGGAGATCTTGGTTCAA 1165
265 sValProGlnAlaGlnProAlaLysProAlaLeuLysSerGlnGluLys 282
|||||
1166 AGTCCCGAGGCGAGCCACCAACCCGACTCAATGACAGAGAGAGA 1215
282 hProValLysLysProGlyAspGlyArgLysValThrPhePheGluPro 298
|:::
1216 CACCATTAAGAAACACAGGGGTGGAAGAAAGTAACTTTTAAACCT 1265
299 SerProGlyAspGluAsnGlyThrSerAsnLysAspGlnGluPheArg 315
|:::
1266 GCGCTTGGGAGTGAATGAGGACTAGTAAAGAGATGAGACTTACAGAT 1315
315 tProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuProMetVal 332
|||||
1316 GCCTATATCAAGTATCAAGAGCTGCTGCGTGGATTTCTCCATGAGTGC 1365
332 rGluValAlaGlnAlaValAlaGlyLysSerGlnGlyHisThrLysAsp 348
|||||
1366 CCGAGGTGCGCCAGCGCTAGAGATTAGTCAAGGACATCACACCAAGAT 1415
349 PheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMet 365
|||||
1416 TTTACAGAGCGCATCTCGAATCTCTCAAGGCGAGTAACTGCCATGAT 1465
365 eAlaArgGluLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIle 382
|||||
1466 AGCCGAGAGTGTGTATGGGGACACTCGGCCACAGCGGACACATTT 1515
382 eLysSerAsnLysSerSerGlyHisValProHisGlyProArgThrArg 398
|||||
1516 TAAAGAAATACATCTCTTACAGCCACGTAACCCATGAGGACTCTCAC 1565
```

```

399 ProSerGluGlnLeuTyrTyrLeuSerArgAlaGlnGlyPheGlnValG1 415
|||||
1566 CCCTCTGAGCAACTGGACATCTTTCCAGAGTCCAGGATTCACAGTTGA 1615
|||||
415 UTYrLyAspPheProLysAsnLysAsnGluCysValSerLeuIleA 432
|||||
1616 ATCAAAAGACTTCCCAAAACAAACAAAGCAATTTGATCTCTATCA 1665
|||||
432 snvSerSerGlnProPheValSerHisGlyIleGlyLysAspVal 448
|||||
1666 ATTGCTCTCTCCAGCCACTGTATCAGCCATGATCGGCAAGGATGTG 1715
|||||
449 GluSerCysHisAspMetAlaIleAlaAsnIleLeuLysLeuLeuSerG1 465
|||||
1716 GACTCGCTGCATATATGGCTGGCTGCAACATCTTAAGTTGCTGTCTGA 1765
|||||
465 ULeuAspGlnGlnSerThrGluMetProArgThrGlyAsnGlyProValS 482
|||||
1766 GTTGGACCAACAAGTACAGATGCCAAGACGAAACGACCAATGT 1815
|||||
482 epiLacysGlyArgCys 487
|||||
1816 CTGTGTGTGGAGGTGC 1832
|||||
seq_name: /SID56/gcgdata/geneseq/geneseqn/NA2000.DAT:293324
seq_documentation_block:
ID 293324 standard; cDNA; 3142 BP.
XX
AC 293324;
XX
DT 04-JUL-2000 (first entry)
XX
DE Human staufen cDNA.
XX
KW Staufen: RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW acquired immune deficiency syndrome; protease; human; ss.
XX
OS Homo sapiens.
XX
FH key Location/Qualifiers
FT CDS 288..1193
FT /tag_a
FT /product= Staufen protein
XX
PN CA2238656-A1.
XX
PD 22-NOV-1999.
XX
PF 22-MAY-1998; 98CA-2238656.
XX
PR 22-MAY-1998; 98CA-2238656.
XX
PA (UYMO-) UNIV MONTREAL.
XX
PI Duchaine T, Wickham L, Cohen EA, Luo M, Desrosellers L,
PI Moulond AJ;
XX
DR WPI: 2000-246924/22.
DR P-PSDB: Y83108.
XX
PT Isolated mammalian or Caenorhabditis elegans stauften proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)
XX
XX Claim 4: Figure 1b; 96pp; English.
XX
XX Staufen is a RNA binding protein which interacts with double stranded
XX RNA and/or rough endoplasmic reticulum. It is involved in mRNA
XX transport and localization. Mammalian SP contain multiple ds
XX RNA-binding domains. Chimeric proteins comprising portions of
XX stauften with proteins with RNase or protease activity can be used to

```

```

CC treat virus infections. The RNase or protease activity of the fusion
CC protein prevents proper maturation of the virus. Mammalian stauften
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hsf (human SP) and
CC MBP/msp (murine SP) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Stauften proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus. This
CC sequence is an alternatively spliced sequence to the one given in
CC 293322.
XX
SO Sequence 3142 BP; 862 A; 778 C; 733 G; 769 T; 0 other:

alignment_scores:
Quality: 2332.00 Length: 489
Ratio: 4.930 Gaps: 1
Percent Similarity: 96.728 Percent Identity: 91.002

alignment_block:
US-09-316-048-11 x 293324
..
Align seg 1/1 to: 293324 from: 1 to: 3142

1 MetTyrLysProValAspProHisSerArgMetGlnSerThrTyrSerTy 17
|||||
309 ATGTATTAAGCGCTGTGATGACCTTACTCTCGATGCACGTCCACTATACATA 358
|||||
17 rGlyMetArgGlyGlyAlaTyrProProArgTyrPheTyrProPhePro 34
|||||
359 CACATGAGAGAGAGGAGTCTTATCCCGAGAGTACTTTACCCATTCCAG 408
|||||
34 aLProProLeuLeuTyrGlnValGluLeuSerValGlyGlyGlnGlnPhe 50
|||||
409 TTCCACCTTACTTCTTATCAAGTGAAGTCTTCTGGAGAGACAGCAATT 458
|||||
51 AsnGlyLysGlyLysMetArgProProValHisAspAlaProAlaAr 67
|||||
459 AATGGCAAGAGAAAGACAGACAGGCTGGCAACAGATGCTCTGCTCCAA 508
|||||
67 gAlaLeuArgThrLeuGlnSerGluProLeuProGluArgLeuGluValA 84
|||||
509 AGCGTTGAGAGATCTCGACAATGAGCCCTGCCAGAGAGCTGAGAGTGA 558
|||||
84 snGlyArgGluAlaGluGluGluAsnLeuAsnLysSerGluIleSerGln 100
|||||
559 ATGGAAGAGATCCGAAGAAATCTCAATTAATCTGAATTAAGTCAA 608
|||||
101 ValPheGluIleAlaLeuLysArgAsnLeuProValAsnPheGluValAl 117
|||||
609 GTGTTTGACATTCACCTTAACGGAACCTGCTGTAAATTTCAAGGTGC 658
|||||
117 aArgGluSerGlyProProHisMetLysAsnPheValThrArgValSerY 134
|||||
659 CCGGAGAGAGTGGCCCAACCCACATGAAGAACTTTGACCAACGTTCCG 708
|||||
134 aGlyGluPheValGlyGluGlyGlyLysSerLysLysIleSerLys 150
|||||
709 TTGGGAGATTGTGGGGGAAGGTGAAGGGGAAGCAAGAAAGATTTCAAAG 758
|||||
151 LysAsnAlaIleAlaArgAlaValLeuGluGlnLeuArgArgLeuProPhe 167
|||||
759 AAAAATGCCCGCATACCTGTTCTTGAGGAGCTGAAGAAGTTACGCCCT 808
|||||
167 uProAlaValGluArgValLysProArgIleLysLysLysSerGlnPro 184
|||||
809 GCCTGAGTTGAACGAGTAAGCCTGAATCAAAAAGAAACAAACCCCA 858
|||||
184 hrcCysLys.....ThrAlaProAspTyrGlyGlnGlyMetAsnProIle 198
|||||
859 TAGTCAAGCCACAGACAGCCCAAGATATGCGCAGGAGTCAATCCGATT 908
|||||

```



```

199 SerATGLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTyr 215
    |||
909 ACCCGACTGGCCAGATCCAGCAGGCAAAAAAGAGAGAGAGAGAGAGTA 958
215 rmetLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMetG 232
    |||
959 CAGGCTCCACAGAGAGCGAGGCTCCCGCGCGCAGGAGATTGTGATGC 1008
232 lValLysValGlyHisHisThrAlaGluGlyValAlaGlyThrAsnLysLys 248
    |||
1009 AGGTGAGAGTTGGAAACCACTGCGAAGAGAGAGCGGCAACCAAGAGAG 1058
249 ValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPheLys 265
    |||
1059 GTGGCCAGACGCAATGACGCGAGAACATGCTGAGATCCTTGTTTCA 1108
265 sValProGlnAlaGlnProAlaLysProAlaLeuLysSerGluGluLysT 282
    |||
1109 AGTCCCGCAGCGGAGCCAGCAAAACCCGACCTCAAGTACAGAGAGAGA 1158
282 hProValLysLysProGlyAspGlyArgLysValThrPhePheGluPro 298
    |||
1159 CACCCATTAAGAAACCGAGGAGTGAAGAAAGTAACCTTTTGAACCT 1208
299 SerProGlyAspGlyAsnGlyThrSerAsnLysAspGluLysPheArgPhe 315
    |||
1209 GGCTCTGGGGATGAATAATGGACTACTATTAAGAGATGAGTTCAGAT 1258
315 tProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuPheMetValP 332
    |||
1259 GCCTTATCTAATCATCAGCAGCTGCTGGAATTTCTCCATGTGTC 1308
332 rGluValAlaGlnAlaValAlaGlyValSerGlnGlyHisHisThrLysAsp 348
    |||
1309 CCGAGGTCGCCAGGCTGTAGAGATTAGTCAGGACATCACACCAAGAAT 1358
349 PheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetIle 365
    |||
1359 TTTACAGAGGACAGCTCGAATCTGCAAGGCCACGCTAACTCCATGAT 1408
365 eAlaArgGluLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleL 382
    |||
1409 ACCCCAGAGATTGTTGATGGGGGACCTGCGCCACAGCCGAGACCATTT 1458
382 euLysSerAsnLysSerSerGlyHisValProHisGlyProArgThrArg 398
    |||
1459 TAAAGATATACATCTCTCAGGCGACGTACCCCATGACCTCTCACAGAG 1508
399 ProSerGluGlnLeuTyrTyrLeuSerArgAlaGlnGlyPheGlnValAl 415
    |||
1509 CCCTCGACCACTGACTATCTTTCCAGATGCCAGGATTCAGAGTTGA 1558
415 uTyrLysAspPheProLysAsnAsnLysAsnGlyCysValSerLeuIleA 432
    |||
1559 ATACAAAGACTTCCCAAAACAAACAAAGCAATTGTATCTCTTATCA 1608
432 sncysSerSerGlnProProLeuValSerHisGlyIleGlyLysAspVal 448
    |||
1609 ATTGCTCCTCTCAGCCACTGTGATCAGCCATGTTATCGGCAAGATGTG 1658
449 GluSerCysHisAspMetAlaAlaLeuAsnIleLeuLysLeuLeuSerG 465
    |||
1659 GAGTCTGCGCATATATGGCTGGCTGACATCTTAAAGTGTGCTGTCTGA 1708
465 uLeuAspGlnGlnSerThrGluMetProArgThrGlyAsnGlyProValS 482
    |||
1709 GTTGGACCAAAAGATACAGAGATGCCAAGAACAGAAACGAGCAATGT 1758
482 eAlaLysGlyLysArgCys 487
    |||
1759 CTGTGTGTGGAGGTGC 1775

```

seq\_name: /SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:293323

seq\_documentation\_block:

ID 293323 standard; cDNA; 3217 BP.

XX 293323;

DT 04-JUL-2000 (first entry)

XX Human stauflen cDNA.

KM Stauflen: RNA binding protein; RNA/DNA hybrid; virus; retrovirus;  
 KM HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;  
 KM acquired immune deficiency syndrome; protease; human; ss.

XX Homo sapiens.

FH Key Location/Qualifiers

FT CDS 363..1853

FT /\*tag= a

FT /\*product= Stauflen

PN CA2238656-A1.

XX 22-NOV-1999.

XX 22-MAY-1998; 98CA-2238656.

XX 22-MAY-1998; 98CA-2238656.

XX (UYMO-) UNIV MONTREAL.

PI Duchaine T, Wickham L, Cohen EA, Luo M, Desgroselliers L;

PI Moulard AJ;

DR WPI; 2000-246924/22.

XX P-PSDB; Y83108.

PT Isolated mammalian or Caenorhabditis elegans stauflen proteins useful  
 for treating retroviral infections especially human immunodeficiency

XX PT virus infections (HIV)

XX PS Claim 4; Figure 1a; 96pp; English.

CC Stauflen is a RNA binding protein which interacts with double stranded  
 CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA  
 CC transport and localization. Mammalian SP contain multiple ds  
 CC RNA-binding domains. Chimeric proteins comprising portions of  
 CC stauflen with proteins with RNase or protease activity can be used to  
 CC treat virus infections. The RNase or protease activity of the fusion  
 CC protein prevents proper maturation of the virus. Mammalian stauflen  
 CC proteins seem to recognise double stranded RNA structure rather than  
 CC any sequence specific position. 2 bacterially-expressed fusion  
 CC proteins used in an RNA-binding assay (his/hsr (human SP) and  
 CC MBP/msp (murine SP) both proteins strongly bound double stranded  
 CC RNA. Both fusion proteins also directly bound labeled ds RNAs and  
 CC RNA/DNA hybrids. Stauflen proteins may therefore be useful in the  
 CC treatment of RNA virus infections, especially those caused by  
 CC retroviruses, in particular human immunodeficiency virus. This  
 CC sequence is an alternatively spliced sequence to the one given in  
 CC 293322.

CC Sequence 3217 BP; 889 A; 796 C; 742 G; 790 T; 0 other;

alignment\_scores:

Quality: 2332.00 Length: 489  
 Ratio: 4.930 Gaps: 1  
 Percent Similarity: 96.728 Percent Identity: 91.002

alignment\_block:

US-09-316-048-11 x 293323

Align seg 1/1 to: 293323 from: 1 to: 3217

```

1 MetTyrLysProValAspProHisSerArgMetGlnSerThrTyrSerTy 17
  |||
384 ATGTATTAACCTGTGACCTTACTCTCGATGCACTGCCATTAACGTA 433
17 rGlyMetArgGlyGlyAlaTyrProProArgTyrPheTyrProPhePro 34
  |||
434 CAACATGAGAGAGAGGCTTATCCCGGAGTACTTTTACCCATTTCGAG 483
34 aLProProLeuLeuTyrGlnValGlyLeuSerValGlyGlnGlnPhe 50
  |||
484 TTCACCTTACTTATCAAGTGAAGTCTTCTGTGGAGACAGCAATTT 533
51 AsnGlyLysGlyLysMetArgProProValLysHisAspAlaProAla 67
  |||
534 AATGGCAAGGAAGACAGACAGGCTGCGAAACAGATGCTGCTCCAA 583
67 gAlaLeuArgThrLeuGlnSerGlnProLeuProGluArgLeuGlnVal 84
  |||
584 ACCGTTGAGGATCTTCGACAAATGAGCCCTGCCAGAGAGGCTGAGGTA 633
84 sncLYArgGlnAlaGlnGluGlnAsnLeuAsnLysSerGlnLysSerGln 100
  |||
634 ATGGAAGAGATCCGAGAAGAAATCTCAATAAATCTGAATAAAGTCAA 683
101 ValPheGlnIleAlaLeuLysArgAsnLeuProValAsnPheGlnVal 117
  |||
684 GTGTTTGAATTTGCACTTAACGGAACTTGTGTAATTTGAGGTGGC 733
117 aArgGlnSerGlyProProHisMetLysAsnPheValThrArgValSerV 134
  |||
734 CCGGAGAGATGGCCGCCACACATGAGAATTTGTGACCAAGTTTCGG 783
134 aGlyGlnPheValGlyGlnGlyLysSerLysLysLysLysLysLys 150
  |||
784 TTGGGAGAGTTTGTGGGGAAGGTGAAGGAAAGCAAGAAATTTCAAG 833
151 LysAsnAlaAlaArgAlaValLeuGlnGlnLeuArgArgLeuProPhe 167
  |||
834 AAAAATGCCGCAATGCTTCTTGAAGAGCTGAAGAAATTCACGCCCT 883
167 uProAlaValGluArgValLysProArgLysLysLysSerGlnProT 184
  |||
884 GCGTCAGTTGAACGAGTAAACCTGATCAAAAAGAAACAAACCA 933
184 hCysLys.....ThrAlaProAspTyrGlyGlnLysMetAsnProIle 198
  |||
934 TAGTCAGAGCCAGACAGACCCAGAAATATGCGAGGAGATCAATCCGAT 983
199 SerArgLeuAlaGlnIleGlnIleAlaLysLysGlnLysGlnProGlu 215
  |||
984 ACCCGACTGGCCCAAGATCCAGCAGGCAAAAGAGAAAGAGACCAAGTA 1033
215 rMetLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMetG 232
  |||
1034 CACGCTCTCCACAGAGCGAGGCTCCCGCGCGAGGAGTTTGATGTC 1083
232 LysValLysValGlyHisHisThrAlaGlnGlyValGlyHisAsnLys 248
  |||
1084 AGGTGAGAGTTGGAACACACTGCGAAGAGAGGCGCAACCAAGAG 1133
249 ValAlaLysArgAsnAlaAlaGlnLysMetLeuGlnLysGlyPheLys 265
  |||
1134 GTGGCCAAACCCAAATGCAACCCAGAAACATGCTGAGATCTTGTTCAA 1183
265 sValProGlnIleGlnProAlaLysProAlaLeuLysSerGlnGluLys 282
  |||
1184 ACTCCGCGAGCGGCGACCCCAACCCGCACTCAAGTCAAGAGAGAGAA 1233
282 hProValLysLysProGlyAspGlyArgLysValThrPhePheGluPro 298
  |||
1234 CACCATTAAGAAACAGGAGATGGAAGAAAGTAACTTTTGTGACCT 1283

```

```

299 SerProGlyAspGlnAsnGlyThrSerAsnLysAspGlnGluPheArg 315
  |||
1284 GGCTCTGGGATGAATAATGAGACTATTAATAAGAGAGATTCAGGAT 1333
315 rProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuProMetValP 332
  |||
1334 GCGTTATCTTAAGTCAATCAGCAGCTGCTCGTGAATTTCTCCATGTC 1383
332 rGlnValAlaGlnAlaValGlyValSerGlnGlyHisHisThrLysAsp 348
  |||
1384 CCGAGCTCCCGAGGCTGTAGGAGTTAGTCAAGAGACATCAACCAAGAT 1433
349 PheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMet 365
  |||
1434 TTACCGAGGCACTCCGATCTCGAATCTGCCAAGGCCAGCTTACTGCC 1483
365 eAlaArgGlnLeuLeuTyrGlyGlyThrSerProThrAlaGlnIleL 382
  |||
1484 AGCCGAGAGTTGTTGTATGGGGGCACTGCCCGACAGCGAGACATTT 1533
382 eLysSerAsnLysSerSerGlyHisValProHisGlyProArgThrArg 398
  |||
1534 TAAAGATTAACATCTTTCAGGCCACGTAACCCATGACCTTCACAGAGA 1583
399 ProSerGlnGlnLeuTyrTyrLeuSerArgAlaGlnGlyPheGlnVal 415
  |||
1584 CCTCTGAGCACTGGACTATCTTTCAGAGTCCAGAGATTCAGAGTTCA 1633
415 uTyrLysAspPheProLysAsnAsnLysAsnGlnLysValSerLeuIle 432
  |||
1634 ATACAAGAGCTTCCCAAAACACAAAGCAAGAAATTTGATCTTATCA 1683
432 sncLysSerSerGlnProProLeuValSerHisGlyIleGlyLysAspVal 448
  |||
1684 ATTGCTCTCTCAGCCACCTCTGATCAGCAATGATCGGCAAGGTGG 1733
449 GluSerCysHisAspMetAlaAlaLeuAsnIleLeuLysLeuSerG 465
  |||
1734 GAGTCTCCCATGATATGCTGCTGCGTGAACATCTTAAATGCTGCTGA 1783
465 uLeuAspGlnGlnSerThrGluMetProArgThrGlyAsnGlyProVal 482
  |||
1784 GTTGAGCAACAAGTACAGAGATGCCAAGAACAGAGAAACGAGCAATGT 1833
482 eArgLysGlyArgCys 487
  |||
1834 CTGTGTGTGGAGGTGC 1850
seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA2000.DAT:293322
seq_documentation_block:
ID 293322 standard; CDNA; 3506 BP.
XX
AC 293322;
XX
DT 04-JUL-2000 (first entry)
XX
DE Human stauflen cDNA.
XX
KW Stauflen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW acquired immune deficiency syndrome; protease; human; ss.
XX
OS Homo sapiens.
XX
FH Key 409..2142 location/Qualifiers
FT CDS /*tag= a
FT /product= Stauflen
XX
PN CA2238656-A1.
XX

```

PD 22-NOV-1999.  
 XX  
 PF 22-MAY-1998; 98CA-2238656.  
 XX  
 PR 22-MAY-1998; 98CA-2238656.  
 XX  
 PA (UYMO-) UNIV MONTREAL.  
 XX  
 PI Duchaine T, Wickham L, Cohen EA, Luo M, Desgroselliers L;  
 PI Moulard AJ;  
 XX  
 DR WPI. 2000-246924/22.  
 DR P-PSDB: Y83023.  
 PT Isolated mammalian or Caenorhabditis elegans staufen proteins useful  
 PT for treating retroviral infections especially human immunodeficiency  
 PT virus infections (HIV)  
 XX  
 XX  
 PS Claim 4; Figure 1a; 96pp; English.

Staufen is a RNA binding protein which interacts with double stranded  
 CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA  
 CC transport and localization. Mammalian SP contain multiple ds  
 CC RNA-binding domains. Chimeric proteins comprising portions of  
 CC staufen with proteins with RNase or protease activity can be used to  
 CC treat virus infections. The RNase or protease activity of the fusion  
 CC protein prevents proper maturation of the virus. Mammalian staufen  
 CC proteins seem to recognise double stranded RNA structure rather than  
 CC any sequence specific position. 2 bacterially-expressed fusion  
 CC proteins used in an RNA-binding assay (his/MSP (human SP) and  
 CC MBP/MSF (murine SP)) both proteins strongly bound double stranded  
 CC RNA. Both fusion proteins also directly bound labeled ds RNAs and  
 CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the  
 CC treatment of RNA virus infections, especially those caused by  
 CC retroviruses, in particular human immunodeficiency virus.

Sequence 3506 BP; 965 A; 875 C; 790 G; 876 T; 0 other;

# alignment\_scores:

Quality: 2332.00 Length: 489  
 Ratio: 4.930 Gaps: 1  
 Percent Similarity: 96.728 Percent Identity: 91.002

## alignment\_block:

us-09-316-048-11 x 293322

Align seg 1/1 to: 293322 from: 1 to: 3506

1 MetTyltSProValAspProHisSerArgMetGlnSerThrTyrSerTyr 17  
 |||||  
 673 ATGATTAAGCGCTGTGACCTTACTCTCGATGAGTCACACCTTAACATA 722  
 17 rGlyMetArgGlyGlyAlaTyrProProArgTyrPheTyrProPhePro 34  
 |||||  
 723 CAACATGAGAGGAGTGTATATCCCGAGGACTTATACCATTTCCAG 772  
 34 alProProLeuLeuTyrGlnValGluLeuSerValGlyGlyGlnPhe 50  
 |||||  
 773 TTCACCTTTACTTATCAAGTGAAGTCTTCGTGGAGGACACCAATT 822  
 51 AsnGlyLysGlyLysMetArgProProValLysHisAspAlaProAlaAr 67  
 |||||  
 823 AATGGCAAGGAAAGACACAGACAGGCTGCGAACAACAGATGCTGCTGCAA 872  
 67 gAlaLeuArgThrLeuGlnSerGluProLeuProGluArgLeuGluValA 84  
 |||||  
 873 AGCGTGTAGATCCTCGAAGATGAGCCCTGCGACAGAGGCTGAGAGGGA 922  
 84 sngLysArgGluAlaGluGluLysLeuSerValGlyGlyGlnSerGln 100  
 |||||  
 923 ATGGAGAGAAATCCGAAAGAAATCTCAATTAATCTGAATAGTCAA 972

101 ValPheGluIleAlaLeuLysArgAsnLeuProValAsnPheGluValAl 117  
 |||||  
 973 GTGTTGAGATTGCACCTTAACGGAACCTTGCTGTAATTCGAGAGTGC 1022  
 117 aArgGluSerGlyProProHisMetLysAsnPheValThrArgValSerV 134  
 |||||  
 1023 CCGGAGAGTGGCCACCACCATGAAGAACCTTGTGACCAAGGCTTCGG 1072  
 134 aLGIyGluPheValGlyGluGlyLysSerLysLysLysLysLys 150  
 |||||  
 1073 TTGGGAGACTTGTGGGGAGGTGAGGAAACACAGAACATTTTCAAG 1122  
 151 LysAsnAlaAlaArgAlaValLeuGluGlnLeuArgArgLeuProProLe 167  
 |||||  
 1123 AAAATAGCGCCATAGCTGTCTTGAAGAGCTGAAGAAGTTACCGCCCT 1172  
 167 uProAlaValGluArgValLysProArgLysLysLysSerGlnProT 184  
 |||||  
 1173 GCGTGCAGTTGACGAGTAAGCCTAGATCAAAAGAAACAAACCA 1222  
 184 hrcysLys.....ThrAlaProAspTyrGlyGlnGlyMetAsnProIle 198  
 |||||  
 1223 TAGTCMAACCACACACACAGCCACATATGGCCAGGGATCAATCCGATT 1272  
 199 SerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTyr 215  
 |||||  
 1273 AGCGGACTGGCCAGATCCACAGGCAAAAGAGAGAGGAGCCAGACTA 1322  
 215 MetLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMetG 232  
 |||||  
 1323 CACCTCTCCACAGAGCGAGGCTCCCGCCGCGCAGGAGTTTGAGAGC 1372  
 232 LysValLysValGlyHisHisThrAlaGluGlyValGlyTyrAsnLysLys 248  
 |||||  
 1373 AGGTGAAGTTGGAACCACTGACAGAGGAGGAGGACCAACAGAGAG 1422  
 249 ValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPhe 265  
 |||||  
 1423 GTGGCCAAAGCCCAATGCGCCGAGACATGCTGAGATCTTGTTGCA 1472  
 265 sValProGlnAlaGlnProAlaLysProAlaLeuLysSerGluGlyST 282  
 |||||  
 1473 AGTCCCGCAGGCGGACCCACCAACCCGACACTCAAGTCAGAGGAGA 1522  
 282 hProValLysLysProGlyAspGlyArgLysValThrPhePheGluPro 298  
 |||||  
 1523 CACCCATTAAGAAACCAAGGAGTGAAGAAAGTAACCTTTTGAACCT 1572  
 299 SerProGlyAspGluAsnGlyThrSerAsnLysAspGluGluPheArgMe 315  
 |||||  
 1573 GGCCTGTGGGATGAATAATGGGACTAGTAATAAGAGATGAGTTCAGAT 1622  
 315 tProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuProMetValP 332  
 |||||  
 1623 GCTTATCTTAATCATAGCAGCTGCTGCTGGAATTCCTCCATGTGGC 1672  
 332 roGluValAlaGlnAlaValGlyValSerGlnGlyHisHisThrLysAsp 348  
 |||||  
 1673 CCGAGAGTGGCCAGGCTGTAGAGTTAGTCAAGGACATCAACCAAAAT 1722  
 349 PheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetL 365  
 |||||  
 1723 TTTACAGAGGAGCTCGCATTCCTGCCAAGGCCAAGGTAACCTGCATGAT 1772  
 365 eAlaArgGluLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleL 382  
 |||||  
 1773 AGCCCGAAGTGTGTATGGGGGACCTCGCCACACACCGACAGACCAATT 1822  
 382 eLysSerAsnLysSerSerGlyHisValProHisGlyProArgThrArg 398  
 |||||  
 1823 TAAAGAAATACATCTTTCAGGCGCACGTAACCCATGACCTCTCAGGGA 1872  
 399 ProSerGluGlnLeuTyrTyrLeuSerArgAlaGlnGlyPheGlnValG 415

```

|||||
1873 CCCTGACGACACTGACTATCTTCCAGAGTCAGGATTCAGGTTGA 1922
415 UYRFLYSPHPHROLYSASNSNLYSASNGLCYSVALSERLEULEA 432
1923 ATACAAAGACTTCCCAAAAACAACAAGCAATTGTACTCTTATCA 1972
432 SNCYSSERSERGLNPROLEUVALSERHISGLYLIEGLYLSASPVAL 448
1973 ATTGCCTCTCTCAGCCACCTGTATCAGCCATGTGTATCGCAAGATGTC 2022
449 GLUSERCYSHISASPMTALAALALEUASNILEULYSLEUUSERTGL 465
2023 GAATCCTGCATGATATGGCTGCGTGAACATCTTAAGTTGCTGTCTGA 2072
465 ULEUASPGINGLNSERTHRLUMETPROARGTHRGYASNGLYPROVALS 482
2073 GTTGGACCAACAACAAAGTACAGATGCCAAGAACGAGAACCAATGT 2122
482 ERLACYSGLYARGCYS 487
2123 CTGTGTGTGGGAGGTGC 2139

seq_name: /SID56/gcdata/geneseq/geneseqn/NA2000.DAT:293327
seq_documentation_block:
ID 293327 standard; cdna; 3260 BP.
XX
AC 293327:
XX
DT 04-JUL-2000 (first entry)
XX
DE Human staufen cdna.
XX
KW Staufen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW acquired immune deficiency syndrome; protease; ; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 409..1311
FT FT /*tag= a
FT FT /product= Staufen protein
XX
CA2238656-A1.
XX
PD 22-NOV-1999.
XX
PF 22-MAY-1998; 98CA-2238656.
XX
PR 22-MAY-1998; 98CA-2238656.
XX
PA (UYMO-) UNIV MONTREAL.
XX
PI Duchaine T, Wickham L, Cohen EA, Luo M, Desrosiers L;
PI Moulard AJ;
XX
DR WPI: 2000-246924/22.
DR P-PSDB: Y83108.
XX
PT Isolated mammalian or Caenorhabditis elegans staufen proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)
XX
PS Claim 4; Figure 1b; 96pp; English.
XX
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC staufen with proteins with RNase or protease activity can be used to
CC treat virus infections. The RNase or protease activity of the fusion

```

```

CC protein prevents proper maturation of the virus. Mammalian staufen
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hsp (human SP) and
CC MBP/msp (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus. This
CC sequence is an alternatively spliced sequence to the one given in
CC 293322.
XX
SQ Sequence 3260 BP; 907 A; 795 C; 759 G; 799 T; 0 other:

```

```

alignment_scores:
Quality: 2316.50 Length: 489
Ratio: 4.908 Gaps: 2
Percent Similarity: 96.524 Percent Identity: 90.798

```

```

alignment_block:
US-09-316-048-11 x 293327 ..

```

```

Align seg 1/1 to: 293327 from: 1 to: 3260

```

```

1 MetTYLSPROVALASPPROHISSEARGMETGLSERTHYRSERY 17
|||||
430 ANGTATAAGCCTGTTGACCTTACTCTCGGATCCAGTCC...TATACATA 476
17 RGLMETARGGLYGLALATYRPROARGTYRPHYRPROPHEROV 34
477 CACATGAGAGAGGAGTGCTTATCCCGAGGTACTTTACCCATTCCAG 526
34 ALPROPROLEULEUTYRGLNVALGLULEUSERYALGLYLINGLNPIE 50
527 TTCACACCTTACTTATCAAGTGGAACTTCTGTGGAGACGCAATTT 576
51 ASNGILYSGLYSMELARGPROPROVALYSHISASPLARPROLAAR 67
577 AATGCCAAGAGAGAGACAGACAGCGCTCCGAAACACAGATCTCTCCAA 626
67 GALALEUARGTHRLEUGLNSERGLUPROLEUPROGLIARGLEUGLVALA 84
627 AGCGTTGAGGATCTCTGACAGATGAGCCCTGCCAGAGAGCTGAGGTGA 676
84 SNGLYARGGLUALGLUGLULASNLEUASNLYSSERGLULESERGIN 100
677 ATGGAAGAGATCCGAAAGAAATCTCAATTAATCTGAATAATGTCAA 726
101 VALPHEGLULLEALALEUULYSARGASNLEUPROVALASNPHEGLUVALA 117
727 GTGTTGAGATTCACCTTAACGGAACCTGCTTAATTTCAGAGTGGC 776
117 AARGLUSERGLYPROPROHISMETLYSASNPIHEVALTHARGVALSERV 134
777 CCGGAGAGTGGCCCAACCCACATGAAGAACTTTGGACCAAGTTTCGG 826
134 ALGLYLUPHEVALGLYLUGLUGLYLYSSERYSLYSILIESERYLS 150
827 TTGGGAGATTGTGGGGGAAAGTGAAGGAAAGCAAGAAATTTAAAG 876
151 LYSASNIALALARGALAYALLEUGLUGLNLNLEUARGLEUPROPROLE 167
877 AAAAATGCCCATACCTGTTCTTGAGAGACTGAACAAGTACCGCCCT 926
167 UPROALAVAGLUARGVALYSPROARGILELYSLYLSYSSERGLNPROT 184
927 GCGTCAGTGAAGAGTAAGCCTGAATCAAAAAGAAACCAACCA 976
184 hrcyslys.....ThrAlaProAspTyrGlyGlnGlyMetAsnProIle 198
977 TACTCAAGCCACGACAGACAGCCAGAAATATGGCCAGGAGATCAATCGAATT 1026

```

```

199 SerArgLeuAlaGlnIleGlnAlaLysLysGluLysGluProGluTyr 215
|||||
1027 AGCGACTGGCCCGATCCAGCAGCAAAAAAGAGAGAGCCAGCAAGTA 1076
215 rMetLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMetG 232
|||||
1077 CACGCTCTCCACAGAGCGAGCCCTCCGCGCCGAGGAGATTGTGATGC 1126
232 lNValLysValGlyHisHisThrAlaGluValGlyTyrHisThrLys 248
|||||
1127 AGGGAAGGTTGGAAACACACCTGCAGAGAGAGGCGCCACCAAGAG 1176
249 ValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPhe 265
|||||
1177 GTGGCCAGAGCGCATGCGCGAGAACATGCTGAGATCCTGTGTTCAA 1226
265 sValProGlnAlaGlnProAlaLysProAlaLysSerGluLysT 282
|||||
1227 AGTCCCGAGCGCGCCACCAAAACCCGCACTCAAGTCAGAGAGAGA 1276
282 hrProValLysLysProGlyAspGlyArgLysValThrPhePheGluPro 298
|||||
1277 CACCATTAAGAAACAGAGGAGTGAAGAAAGAACCTTTTGTGAACCT 1326
299 SerProGlyAspGluAsnGlyThrSerAsnLysAspGluGluPheArgMe 315
|||||
1327 GGCCTGGGGATGAATAATGGAGCTAGTAATTAAGAGATGAGTTCAGAT 1376
315 tProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuProMetValP 332
|||||
1377 GCCTTAATCTAAGTCATCAGCAGCTGCTGCGAATTCCTCCATGTGTC 1426
332 roGluValAlaGlnAlaValGlyValSerGlnGlyHisHisThrLysasp 348
|||||
1427 CCGAGGTGCGCCAGCTGTAGAGTTAGTCAAGACATCACACCAAGAT 1476
349 PheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetI 365
|||||
1477 TTTCAGAGGGAGCTCGGAATCCGCCAGGCCAGGTAATGCCATGAT 1526
365 eAlaArgGluLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIle 382
|||||
1527 AGCCCGAGAGTTGTTGATGGGGCACCCTCCGCCACAGCCGAGACATTT 1576
382 euLysSerAsnIleSerSerGlyHisValProHisGlyProArgThrArg 398
|||||
1577 TAAAGATTAACATCTCTTCAGGCGACGACCCCATGGACCTCTCAGAGA 1626
399 ProSerGluGlnLeuTyrTyrLeuSerArgAlaGlnGlyPheGlnValG 415
|||||
1627 CCCCTGAGCAACTGCACTATCTTCCAGAGTCCAGGATTCAGGTTGA 1676
415 uTyrLysAspPheProLysAsnAsnLysAsnGlyCysValSerLeuIleA 432
|||||
1677 ATACAAGACTTCCCAAAAACAACAGAACGAATTTCTATCTCTTATCA 1726
432 snCysSerSerGlnProProLeuValSerHisGlyIleGlyLysAspVal 448
|||||
1727 ATTGCTCTCTCAGCCACTCTGTATCAGCCATGGATATGCGCAAGGATGTG 1776
449 GluSerCysHisAspMetAlaAlaLeuAsnIleLeuLysLeuLeuSerG 465
|||||
1777 GAGTCCTGCCCATGATATGGCTGCCGTGAACATCTTAAAGTTGCTGTGA 1826
465 uLeuAspGlnGlnSerThrGluMetProArgThrGlyAsnGlyProValS 482
|||||
1827 GTTGACCAACAAAGTACAGAGATGCCAAGAAACAGAAACGAGCAACATGT 1876
482 eAlaGlyGlyArgCys 487
|||||
1877 CTGTGTGTGGAGGTGC 1893

```

seq\_name: /SID56/gcgdata/geneseq/geneseqn/NA1999.DAT:X39991

```

seq_documentation_block:
ID      X39991 standard; DNA; 1127 BP.
XX
AC      X39991;
XX
DT      02-JUL-1999 (first entry)
XX
DE      Prostate cancer associated gene.
XX
KW      Cancer associated antigen; diagnosis; research; treatment; human;
KW      breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW      prostate cancer; ss.
XX
OS      Homo sapiens.
XX
PN      WO904265-A2.
XX
PD      28-JAN-1999.
XX
PF      15-JUL-1998; 98WO-US14679.
XX
PR      22-JUN-1998; 98US-0102322.
PR      17-JUL-1997; 97US-0896164.
PR      10-OCT-1997; 97US-0061599.
PR      10-OCT-1997; 97US-0061765.
PR      10-OCT-1997; 97US-0948705.
PR      11-OCT-1997; 97GB-0021697.
XX
PA      (LUDW-) LUDWIG INST CANCER RES.
XX
PI      Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI      Pfeundschnun M, Sahin U, Scanlan MJ, Stockert E;
PI      Tureci O.
XX
DR      WPI: 1999-132448/11.
XX
PT      New isolated sera associated nucleic acids and polypeptides -
PT      isolated using sera from cancer patients, used to develop products
PT      for the diagnosis, monitoring or treatment of cancers
XX
PS      Claim 67: Page 625-626; 787pp: English.
XX
CC      The invention relates to a method for diagnosing a disorder characterised
CC      by expression of a human cancer associated antigen precursor coded for by
CC      a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC      biological sample isolated from a subject with an agent that specifically
CC      binds to the NAM, an expression product or a fragment of an expression
CC      product complexed with an HLA molecule; and (b) determining the
CC      interaction between the agent and the NAM or the expression product as a
CC      determination of the disorder. The products and methods can be used in
CC      the diagnosis, monitoring, research, or treatment of conditions
CC      characterised by the expression of various cancer associated antigens.
CC      The invention provides nucleic acid sequences and encoded polypeptides
CC      which are cancer associated antigen precursors expressed in human breast
CC      cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC      lung cancer.
XX
SQ      Sequence 1127 BP; 349 A; 265 C; 307 G; 203 T; 3 other;
XX

```

Alignment\_scores:

Quality:	1710.00	Length:	375
Ratio:	4.763	Gaps:	2
Percent Similarity:	95.733	Percent Identity:	89.067

Alignment\_block:

US-09-316-048-11 x X39991 ..

Align seg 1/1 to: X39991 from: 1 to: 1127

```

47 GlyGlnGlnPheAsnGlyLysGlyLysMetArgProProValLysHisAs 63
|||||

```

```

3  GGACACCAATTTAATGGCAAGGAAAGACAGAGAGCTGCGAAGACAGA 52
63  PALAPROALAArgAlaLeuArgThrLeuGlnSerGluProLeuProGluA 80
53  TGGTGTGCGCAAGCGTTGAGATCTCGCAGAAATGAGCCCTGCCAGAGA 102
80  rglLeuGluValaInglArgGluAlaGluGluLysLeuAsnLysSer 96
103  GGCTGGAGGATGGAAGAGATCCGAGAGAAATCATCATTAATCT 152
97  GluILSerGlnValPheGluILLeuLeuLysArgAsnLeuProValas 113
153  GAATTAAGTCAACTGTTGAGATTGACCTTAACGGAACTGCTGGAA 202
113  nphGluValAlaArgGlnSerGlyProProHISmetLysAsnPhenAla 130
203  TTTCGAGGTGGCCGGGAGAGTGGCCACCACATGAAACACTTGTGA 252
130  hrArgValSerValGlyGluPheValGlyGluGlyLysSerLys 146
253  CCAAGGTTTCGTTGGGAGATTGTGGGGAGAGCTGAGGGAAGCAAG 302
147  LysILSerLysLysAsnAlaAlaArgAlaValLeuGluGlnLeuArg 163
303  AAGATTTCAAAGAAAATGCCCATAGCTGTTCTGAGAGCTGAAGAA 352
163  glLeuProProLeuProAlaValGluArgValLysProArgILLeuLysL 180
353  GTTACGCCCCCTGCTGCACTGACAGAGTAAAGCCATGATCAAAAGA 402
180  ysSerGlnProThrCysLys.....ThrAlaProAspTyrGlyGlnGly 194
403  AACCAAAACCCATGATGACAGCCACAGCAAGCCAGATATGCGCAGGG 452
195  MetAsnProILeserArgLeuAlaGlnILLeuGlnAlaLysLysGlu 211
453  ATCAATTCGATTAGCCGACTGGCCCATCTCACAGGCAAAAAAGAGAA 502
211  sglUProGluTyrMetLeuLeuThrGluArgGlyLeuProArgArg 228
503  GAGGCCAGAGTACACGCTCTCCACAGAGCGAGGCTCCGCGCGCAGG 552
228  LuPheValMetGlnValLysValGlyHISThrAlaGluGlyLys 244
553  AGTTTGATGACGAGTGAAGCTTGAACACCACTGACAGAGAGAGCGGC 602
245  ThrAsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGlu 261
603  ACCAACAAGAGGTGGCCCAAGCGCAATGACCGAGAGAACTGTGGAGAT 652
261  eLeuGlyPheLysValProGlnAlaGlnProAlaLysProAlaLeuLys 278
653  CCTTGTTTCAAAAGTCCCGAGCGCAGCCACCAACCGCAGCTCAAGT 702
278  erGluGluLysThrProValLysLysProGlyLysPglYArgLysValThr 294
703  CAGAGAGAGAACACCCATTAAGAAACACAGGGATGAGAAAGAAATACC 752
295  PhePheGluProSerProGlyAspGluAsnGlyThrSerAsnLysAsp 311
753  TTTTGTGAACCTGGCTCTGGGGATGAATAAGCTAGTAATAAGAGA 802
311  uGluPheAlaGlyMetProTyrLeuSerHISGlnGlnLeuProAlaGlyIL 328
803  TGAAGTTCAAGATGCTTATCTAAGTCATCAGCAGCTGCTGCTGGAAATTC 852
328  euProMetValProGluValAlaGlnAlaValGlyLysSerGlnGlyHIS 344
853  TTCCCATGTGTCGCCGAGGTGCGCCAGGCTGTAGAGTGAAGCAAGCAT 902
345  HISThrLysAspPheThrArgAlaAlaProAsnProAlaLysAlaThrY 361
903  CACACCAAGATTTTACAGGGGAGCTCGAAATCTGCCAAGGCCACGCGT 952

```

```

361  lThrAlaMetILeAlaArgGluLeuLeuTyrGlyLysThrSerProThrA 378
953  AACTGCCATGATAGCCCGAGAGTGTGTATGGGGCAGCTGCCACAG 1002
378  la.GluThrILLeuLysSerAsnILeserSerGlyHISValProHISgl 394
1003  CCCGAGACCATTTTAAAGATATACATCTTTCAGGCACAGTACCCATGG 1052
394  yProArgThrArgProSerGluGlnLeuTyrTyrLeuSerArgAlaGln 411
1053  ACCTTCACAGAGACCCCTMTGAGCACTGACATCTTTCAGAGTCCAG 1102
411  lPheGlnValGluTyrLysAsp 418
1103  GATTNCAGCTTGATATCCAAAGAC 1125

seq_name: /SID56/gcgdata/geneseq/geneseqn/NA1999.DAT:X39992
seq_documentation_block:
ID      X39992 standard; DNA: 773 BP.
XX
AC      X39992;
XX
DT      02-JUL-1999 (first entry)
XX
DE      Prostate cancer associated gene.
XX
KW      Cancer associated antigen; diagnosis; research; treatment; human;
KW      breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW      prostate cancer; ss.
XX
OS      Homo sapiens.
XX
PN      W09904265-A2.
XX
PD      28-JAN-1999.
XX
PF      15-JUL-1998; 98WO-US14679.
XX
PR      22-JUN-1998; 98US-0102322.
PR      17-JUL-1997; 97US-0896164.
PR      10-OCT-1997; 97US-0061599.
PR      10-OCT-1997; 97US-0061765.
PR      10-OCT-1997; 97US-0948705.
PR      11-OCT-1997; 97GB-0021697.
XX
PA      (LUDW-) LUDWIG INST CANCER RES.
XX
PI      Chen Y, Gout I, Gure A, O'Hare M, Ohta Y, Old LJ;
PI      Pfeundschnuh M, Sahlin U, Scanlan MJ, Stockert E;
PI      Tureci O;
XX
DR      WPI: 1999-132448/11.
XX
PT      New isolated cancer associated nucleic acids and polypeptides -
PT      isolated using sera from cancer patients, used to develop products
PT      for the diagnosis, monitoring or treatment of cancers
XX
PS      Claim 67; Page 626; 787pp; English.
XX
CC      The invention relates to a method for diagnosing a disorder characterised
CC      by expression of a human cancer associated antigen precursor coded for by
CC      a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC      biological sample isolated from a subject with an agent that specifically
CC      binds to the NAM, an expression product or a fragment of an expression
CC      product complexed with an HLA molecule; and (b) determining the
CC      interaction between the agent and the NAM or the expression product as a
CC      determination of the disorder. The products and methods can be used in
CC      the diagnosis, monitoring, research or treatment of conditions
CC      characterised by the expression of various cancer associated antigens.
CC      The invention provides nucleic acid sequences and encoded polypeptides
CC      which are cancer associated antigen precursors expressed in human breast

```

CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
 CC lung cancer.  
 XX  
 SQ Sequence 773 BP; 233 A; 204 C; 199 G; 134 T; 3 other;

alignment\_scores:  
 Quality: 1183.00 Length: 257  
 Ratio: 4.789 Gaps: 2  
 Percent Similarity: 96.109 Percent Identity: 89.494

alignment\_block:  
 US-09-316-048-11 x X39992

Align seg 1/1 to: X39992 from: 1 to: 773

```

165 ProProleuProAlaValGluArgValLysProArgLysLysLysSe 181
      |||||||.....
3  CCCCCCTGCTGCTGAGTGAACGATTAACCTAAGATCAAAAAGAAAC 52
      |||||||.....
181 rglnProThrCysLys.....ThralProAspTyrGlyGlnGlyMeLa 196
      |||.....
53 AAAACCATATGTCACAGCCACAGACACCCAGATATGCGCAGGGATCA 102
      |||.....
196 snProIleSerArgLeuAlaGlnIleGlnAlaLysLysGlyLysGlu 212
      |||||||.....
103 ATCCGATTACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 152
      |||||||.....
213 ProGluTyrMetLeuLeuThrGluArgGlyLeuProArgArgGluPh 229
      |||||||.....
153 CCAGAGTACACGCTCTCCACAGAGCGAGCTCCGCGCGCAGGGAGTT 202
      |||||||.....
229 eValMetGlnValLysValGlyHisThrAlaGluGlyValGlyThra 246
      |||||||.....
203 TGGAATGACAGTGAAGTTGGAACCACTGCAAGAGAAAGGGACCA 252
      |||||||.....
246 snLysLysValAlaLysArgAsnAlaLysLysLysLysLysLysL 262
      |||||||.....
253 ACAGAGAGTGGCCAGGCAATGCAAGCAAGCAATGCTGAGATCTT 302
      |||||||.....
263 GlyPheLysValProGlnAlaGlnProAlaLysProAlaLeuLysSerG 279
      |||||||.....
303 GGTTCCTAAAGTCCCGCAGCGCAGCCACCAACCCCTCCTCAAGTCA 352
      |||||||.....
279 uGlyLysThrProValLysLysProGlyAspGlyArgLysValThrPhe 296
      |||||||.....
353 GGAAGAACACCCATTAAGAAACAGGGGATGGAAGAAAGTAACTTTT 402
      |||||||.....
296 heGluProSerProGlyAspGluAsnGlyThrSerAsnLysAspGlu 312
      |||||||.....
403 TTGAACCTGCTGCTGCGGATGAAATGGACTAGTAAATAAGAGATGAG 452
      |||||||.....
313 PheArgMetProTyrLeuSerHisGlnGlnLeuProAlaGlyLysLeuP 329
      |||||||.....
453 TTCAGATGCTTATCTAATCATCAGCAGCTGCTGCTGCTGCTGCTGCT 502
      |||||||.....
329 oMetValProGluValAlaGlnAlaValGlyValSerGlyHisHisT 346
      |||||||.....
503 CATTGTCCTCCGAGGTGCGCCAGGCTGAGAGTTAGTCAAGGACATCA 552
      |||||||.....
346 hrLysAspPheThrArgAlaAlaProAsnProAlaLysAlaThrValThr 362
      |||||||.....
553 CCAAAAGTTTACCAAGGACGCTCCGAATCTCCCAAGGCCACGCTAACT 602
      |||||||.....
363 AlaMetIleAlaArgGluLeuLeuTyrGlyLysThrSerProThrAla 379
      |||||||.....
603 GCCATGATACCCGAGAGTGTGTATGGGGGACCTCGCCACAGCCCG 652
      |||||||.....
379 luhTrIleLeuLysSerAsnLysSerSerGlyHisValProHisGlyPro 395
      |||||||.....
653 AGACCATTTTAAAGAAATACATCTCTTCAGGCCACGTAACCCATGACCT 702
      |||||||.....
396 ArgThrArgProSerGluGlnLeuTyrTyrLeuSerArgAlaGlnGlyPh 412

```

```

|||||.....
703 CTCACGAGACCTTWTGAGCAACTGACTATCTTNCAGAGTCAGGAT 752
      |||||||.....
412 eGlnValGluTyrLysAsp 418
      |||||||.....
753 NCAGTTGAATACCAAGAC 771

```

seq\_name: /SID56/gcgsdata/geneseq/geneseqn/NA1999.DAT.X40000

seq\_documentation\_block:  
 ID X40000 standard; DNA; 844 BP.

AC X40000;  
 XX  
 XX  
 DT 02-JUL-1999 (first entry)  
 XX

DE Prostate cancer associated gene.

XX Cancer associated antigen; diagnosis; research; treatment; human;  
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
 KW prostate cancer; ss.

OS Homo sapiens.

XX  
 PN W09904265-A2.

XX  
 PD 28-JAN-1999.

XX  
 PE 15-JUL-1998; 98WO-US14679.

XX  
 PR 22-JUN-1998; 98US-0102322.

XX  
 PR 17-JUL-1997; 97US-0896164.

XX  
 PR 10-OCT-1997; 97US-0061599.

XX  
 PR 10-OCT-1997; 97US-0061765.

XX  
 PR 10-OCT-1997; 97US-0948705.

XX  
 PR 11-OCT-1997; 97GB-0021697.

XX  
 PA (LUDW-) LUDWIG INST CANCER RES.

XX  
 PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;

XX  
 PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;

XX  
 PI Tureci O;

XX  
 DR WPI; 1999-132448/11.

XX  
 PT New isolated cancer associated nucleic acids and polypeptides -

XX  
 PT isolated using sera from cancer patients, used to develop products

XX  
 PT for the diagnosis, monitoring or treatment of cancers

XX  
 PS Claim 67; Page 629; 787pp; English.

XX The invention relates to a method for diagnosing a disorder characterised  
 CC by expression of a human cancer associated antigen precursor coded for by  
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
 CC biological sample isolated from a subject with an agent that specifically  
 CC binds to the NAM, an expression product or a fragment of an expression  
 CC product complexed with an HLA molecule; and (b) determining the  
 CC interaction between the agent and the NAM or the expression product as a  
 CC determination of the disorder. The products and methods can be used in  
 CC the diagnosis, monitoring, research, or treatment of conditions  
 CC characterised by the expression of various cancer associated antigens.  
 CC The invention provides nucleic acid sequences and encoded polypeptides  
 CC which are cancer associated antigen precursors expressed in human breast  
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
 CC lung cancer.  
 XX

SQ Sequence 844 BP; 251 A; 196 C; 234 G; 156 T; 7 other;

alignment\_scores:  
 Quality: 1164.50 Length: 281  
 Ratio: 4.531 Gaps: 3  
 Percent Similarity: 91.459 Percent Identity: 83.274





```

47 GlyGlnGlnPheasnGlyLysGlyLysMetArgProProValLysHisAs 63
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
   3 GGACGACAAATTTAATGGCAAAAGCAAAAGCAAGCGCTGGCAAAACAGA 52
63 PAlaProAlaArgAlaLeuArgThrLeuGlnSerGluProLeuProGluA 80
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
   53 TGGCTGCGCAAAAGCTTGAAGATCTCTGCAAGAAAGAGCCCTGCCAAGA 102
80 rLeuGluValAsnGlyArgGluValGluGluValAsnLeuAsnLysSer 96
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
   103 GCGCGAGGAGTAATGGAAGAAATCCAGACAGAAATCTCAATAATCT 152
97 GluLeuSerGlnValPheGluLeuLeuLeuLysArgAsnLeuProValAs 113
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
   153 GAATAATAGTCAAGTGTGAGATTCACCTTAACGGAACCTTGCTGGAA 202
113 nPheGluValAlaArgGluSerGlyProProHisMetLysAsnPheValT 130
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
   203 TTTTCGAGGTGGCCGGGAGAGTGGCCACCCACATGAGAACTTTGTGA 252
130 hrArgValSerValGlyLupheValGlyGluGluGluGluLysSerLys 146
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
   253 CCAAGGTTTGGTTGGGAGTTTGTGGGGAAGGTGAAGGGAAGAAAGCAAG 302
147 LysLeuSerLysAsnAlaAlaArgAlaValLeuGluGluLeuArgAr 163
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
   303 AAGATTTCAAAGAAATGCCGCCATAGCTGTCTTGAGGAGCTGAGAA 352
163 gLeuProProLeuProAlaValGluArgValLysProArgLLeuLysL 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
   353 GTTACCCGCCCTGCTCAGTTGAACAGTAAAGCTAAGATCAAAAAGA 402
180 ySerGlnProThrCysLys..... ThrAlaProAspTyrGlyGly 194
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
   403 AAACAAACCATATGATCAAGCCACAGACAGCCAGCAATATGCGCAGGG 452
195 MetAsnProLeuSerArgLeuAlaGlnLeuGlnAlaLysLysGly 211
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
   453 ATCATCTCGATTACCGCAGCTGCCACATCCAGACGCAAAAGGAGAA 502
211 sGluProGluTyrMetLeuLeuThrGluArgGlyLeuProArgArg 228
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
   503 GGACCCAGAGTACAGCTCTCCACAGAGCGAGCGCTNCCGCGCGCAGGG 552
228 luePheValMetGlnValLysValGlyHisThrAlaGluGlyValG 244
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
   553 AGTTTGTATGTCAGGTGAAGGTGAAACCACTTGCAGAAAGAACCGG 602
244 yThrAsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuG 261
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
   603 CACCAACAGAAAGTGGCCAAAGCCCAATGCACCCGAGAACATGCTGAGA 652
261 leLeuGlyPheLysValProGlnAlaGlnProAlaLysProAlaLeu 277
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
   653 TCCTTGTTCATAAAGTCCCGCANGCCAGCCCAACCAACCCGAGACTNA 702
277 sSerGluGluLysThrProValLysLysProGlyAspGlyArgLysValT 294
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
   703 GTCCAGAGAGAAAGCCCATTAAGAAACCAAGGATGGAAGAAAGTAA 752
294 hrPhePheGluProSerProGlyAspGluAsnGlyThrSerAsnLysAsp 310
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
   753 CNTTTGAANCTGGCTNTTGGGA...TTAAATGGCGCTTGTANTRAAAGAG 799
311 GluLupheArgMetProTyrLeuSerHisGlnLeuProAlaGlyI 327
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
   800 GATGATGTTACAGATGNCATATCAATCAATNAACACTTGCT. GCTGGA 848
327 eleuProMetValProGluValAlaGlnAlaValGlyValSerGlnGly 344
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
   849 TCTTTCCATGGGGCCCGAGGTCNCCAGCTTTTGAAGTATATCANAGCC 898
344 lshlThrLysAspPheThrArgAlaAlaPro 354

```

```

seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA2000.DAT:z79981
seq_documentation_block:
ID 279981 standard: cDNA; 660 BP.
XX
AC 279981;
XX
DT 07-APR-2000 (first entry)
DE
DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:65.
XX
XX Human: gene expression product; diagnosis: tumour; colon cancer;
KW colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
KW hyperplasia; ds.
XX
OS Homo sapiens.
XX
PN M09964576-A2.
XX
PD 16-DEC-1999.
XX
PF 09-JUN-1999; 99NC-IB01062.
XX
PR 10-JUN-1998; 98US-0088801.
XX
PA (FARB ) BAYER CORP.
XX
PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;
XX
DR WPI: 2000-087220/07.
XX
PT Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer
XX
XX Claim 1; Page 158; 469pp; English.
XX
CC 279917 to 280766 represent double stranded cDNA clones isolated from the
CC human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g. carcinomas
CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used
CC for developing agents for the diagnosis and treatment of disorders
CC involving unwanted cell proliferation, such as neoplasia, dysplasia or
CC hyperplasia.
XX
SO Sequence 660 BP; 133 A; 172 C; 159 G; 174 T; 22 other;

alignment_scores:
Quality: 678.50 Length: 217
Ratio: 3.791 Gaps: 3
Percent Similarity: 82.488 Percent Identity: 69.124

alignment_block:
US-09-316-048-11 x 279981/rev ..
Align seg 1/1 to reverse of: 279981 from: 1 to: 660
176 ArgIleLysLysSerGlnProThrCysLysThrAlaProAspTyrG 192
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
644 CGGATACATTNNCCNAGAACCAAGTTGCGCATGTTNCCCAANGTGTGC 595
192 yGlnGly MetAsnProLeuSerArgLeuAlaGlnLeuGlnAlaLys 208
::: ::::: |||

```

```

594 CGAAGTTGATTCNNAGTAGCGCCNCCAGTGNGTGGANTTCGCCNTA 545
209 LysGluLysGluProGluTyrMetLeuLeuThrGluArgGlyLeuProAr 225
    ::::: ::::: ::::: :::::
544 AGCGTGTCCGCCCAAGTAGACAGTTCTTCANCAAGCAGCCCTCCGCGN 455
225 gATgATgGgLU. PheValMetGlnValLysValGlyHisThrAlaGlu 241
    ::::: ::::: ::::: :::::
494 GCGNAGGAGAGTTGTATNATCAGGTNAAGCTTAGAACCCCATTTGCANAA 445
242 GLyValGlyThrSnlLysValAlaLysArgAsnAlaLaGluAsnMe 258
    ::::: ::::: ::::: :::::
444 GGACG. GCGCACACCAAGAGG...TGGCCAAGGCAATGACGCGAGANCAT 399
258 tLeuGluLLeuGlyPheLysValProGlnAlaGlnProAlaLysProA 275
    ::::: ::::: ::::: :::::
398 GTTGAAGATCTCGGTTTCAAAAGTCCCGCAGCGCCAGCCACC. AAACCGG 350
275 lAlLysSerGluGluLysThrProValLysLysProGlyAspGlyArg 291
    ::::: ::::: ::::: :::::
349 CACTCAAGTCAGAGAGAGAACACCCCTAAAGAAACACGAGGATGAGAGA 300
292 LysValThrPhePheGluProSerProGlyAspGluAsnGlyThrSerAs 308
    ::::: ::::: ::::: :::::
299 AAGTAACTCTTTTGAACCTGCTGCGGATGAAATGGACTAGTAA 250
308 nLysAspGluGluPheArgMetProTyrLeuSerHisGlnGluLeuProA 325
    ::::: ::::: ::::: :::::
249 TAAAGAGAGATGAGTTCAGATGCTTATCTAAGTCATCAGACAGCTGCTG 200
325 lAcGlyLeuProMetValProGluValAlaGlnAlaValGlyValSer 341
    ::::: ::::: ::::: :::::
199 CTGGAATTCCTCCATGCTGCCGAGGTCGCCACGCTGTAGAGATTAGT 150
342 GlnGlyHisHisThrLysAspPheThrArgAlaAlaProAsnProAlaLy 358
    ::::: ::::: ::::: :::::
149 CAAGGACATCACACCAAGATTTTACAGGCGACGCTCGAATCCTGCGCAA 100
358 sAlaThrValThrAlaMetIleAlaArgGluLeuLeuTyrGlyThrS 375
    ::::: ::::: ::::: :::::
99 GGGCAGCGTAGCTGCCATGATAGCCCGAGAGTGTGTATGGGGGCACCT 50
375 ePrProThrAlaGluThrIleLeuLysSerAsnLysSerGlyHis 390
    ::::: ::::: ::::: :::::
49 CGCCACAGCCGAGACCATTTTAAAGATTAACATCTCTCAGGCCAC 3
seq_name: /SID56/gcgdata/geneseq/geneseqn/NA2000.DAT:280265
seq_documentation_block:
ID 280269 standard; cDNA; 727 BP.
AC 280269;
XX
DT 07-APR-2000 (first entry)
XX
DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:353.
XX
KW Human; gene expression product; diagnosis; tumour; colon cancer;
    colorectal adenocarcinoma; cell line SW480; cell proliferation;
    cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
    hyperplasia; ds.
XX
OS Homo sapiens.
XX
PN WO9964576-A2.
XX
PD 16-DEC-1999.
XX
PE 09-JUN-1999; 99WO-1B01062.
XX
PR 10-JUN-1998; 98US-0088801.
XX
PA (FARB ) BAYER CORP.

```

```

XX Endege WO, Steilmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catlino TJ, Dertl A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;
XX WPI: 2000-087220/07.
XX
PT Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer
XX
PS Claim 15; Page 273; 469pp; English.
XX
CC Z79917 to 280766 represent double stranded cDNA clones isolated from the
CC human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g. carcinomas
CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used
CC for developing agents for the diagnosis and treatment of disorders
CC involving unwanted cell proliferation, such as neoplasia, dysplasia or
CC hyperplasia.
XX
SQ Sequence 727 BP; 194 A; 153 C; 191 G; 149 T; 40 other;

```

```

alignment_scores:
    Quality: 633.00      Length: 166
    Ratio: 4.336        Gaps: 3
Percent Similarity: 87.952 Percent Identity: 77.711

```

alignment\_block:

us-09-316-048-11 x 280269

Align seq 1/1 to: 280269 from: 1 to: 727

```

28 TyrPheTyrProPheProValProProLeuLeuTyrGlnValGluLeuSe 44
    ::::: ::::: ::::: :::::
3 TACTTTTACCACATTTCCAGTTCACGCTTACTTATCAAGTGAAGACTTTC 52
44 rValGlyGlyGlnGlnPheAsnGlyLysGlyLysMetArgProProVal 61
    ::::: ::::: ::::: :::::
53 TGTGGAGAGCAGACCAATTTAATGGCAAGGAAAGACAGACAGCGTCGA 102
61 yHisAspAlaProAlaArgAlaLeuArgThrLeuGlnSerGluProLeu 77
    ::::: ::::: ::::: :::::
103 AACACGATCTGCTGCCAAGCGTTGAGGATCTTGCAGAGATGAGCCCTG 152
78 ProGluArgLeuGluValAsnGlyArgGluAlaGluGluAsnLeuAs 94
    ::::: ::::: ::::: :::::
153 CCAGAGAGCGGTGAGGTGAATGGAAGAGATCCGAGAGCAAAATATCTCA 202
94 nLysSerGluLLeSerGlnValPheGluLLeAlaLeuLysArgAsnLeuP 111
    ::::: ::::: ::::: :::::
203 TAAATCTGAATTAAGTCAAGTGTGATGATTCATTAACGGAACCTTGC 252
111 roValAsnPheGluValAlaArgGluSerGlyProProHisMetLysAsn 127
    ::::: ::::: ::::: :::::
253 CTGTGAATTTCCAGTGTGCCCGGAGAGTGGCCACCCACATGAAGAAC 302
128 PheValThrArgValSerValGlyGluPheValGlyGlyGlyGly 144
    ::::: ::::: ::::: :::::
303 TTTGTGACCAAGGTTTGGTTGGGAGTTTGGGGAAGGGAAGGAA 352
144 sSerLysLysLLeSerLysLysAsnAlaAlaArgAlaValLeuGluGln 160
    ::::: ::::: ::::: :::::
353 AACCAAGAGATTTCAAGAAATGCCCGCATGACTGTCTTGAAGAG 402
161 LeuArgArgLeuProProLeuProAlaVal...GluArgValLysProAr 176
    ::::: ::::: ::::: :::::
403 CTGAGAGAAAGTACCGCCCTGCTGTGATTTGAGCAGCAGATTAAGGCTTAN 452

```



```

277 sserGIuLyuStHrProValIySAspGIyAArgLySValT 294
676 .....GCCA 679
294 hrPhepHeGluProSerProGlyAAspGluAsnGlyThrSerAsnLySAsp 310
680 AATTAGTATATTCTCCA.....GAGACCACCATTTCT..... 714
311 GLuGluPheArgMetProTyLeuSerHisGluGlnLeuProAlaGlyI 327
714 ..... 714
327 eLeuProMetValProGluValAlaGlnAlaValGlyValSerGlyGly 343
715 .....TTAACAAATGCTAGTACGACATCTTTAGAT 745
344 .....HisHisThrLySAspPheThrArgAlaIaIaProAsnProAla 357
746 GTACTTGCAATTCCTTGAGGAAT..... 768
358 LysAlaThrValThrAlaMetIleAlaArgGluLeuLeuTyGlyGlyTh 374
768 ..... 768
374 rSerProThrAlaGlu...ThrIleLeuLySAsnIleSerSerGlyH 390
769 .TCCTCGTGAAGATCACTACTGAAAGAACCTTCCTAGT.... 813
390 lSValProHisGlyProArgThrArgProSerGluGlnLeuTyGlyLeu 406
814 .....ATTCCAAATACAGATTACATCCAGCTGCTTAGTGAAT 852
407 SerArgAlaGlnGlyPheGlnValGluTyGlyLySAspPheProLyS.... 421
853 GCCAAGAACAAAGTTTAAATATACATTTTGGATATAGATGACCTGAG 902
422 ...AsnAsnLySAsnGluCySValSerLeuIleAsnCySserSerGlnP 437
903 CCCCAATGACAAATATCATGT.....CTTGCTGAACGTCTCCACGAGCC 946
437 rOProLeuValSerHisGlyIleGlyLySAspValGluSerCySHisAsp 453
947 CCATCAACAGCTCTGTCAATGCTCCGGTATCTCCTGGCAATGACAAAGCT 996
454 MetAlaIaIaLeuAsnIleLeuLySLeuLeu..... 463
997 GATGCAGCTCACAAATGCTTGCATTTTAAAGATATATAGCAGAAAGAA 1046
464 .....SerGluLeuAspGlnGlnSerThrGluMetProArgThrG 477
1047 GTAATCTGACCACTTAAAAAATCTTTCAGTAGACACATAAAAAGTTCC 1096
477 LysAsnGlyPro 480
1097 CCTCTGGCCCC 1107

seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1999.DAT:X03457
seq_documentation_block:
ID X03457 standard; cDNA; 1608 BP.
XX
AC X03457;
XX
DT 07-APR-1999 (first entry)
XX
DE cDNA encoding a double-stranded RNA-binding protein.
XX
KW Human; double-stranded RNA-binding protein; BINDR; Incyte clone 620438;
KW adenocarcinoma; leukemia; lymphoma; melanoma; myeloma; sarcoma;
KW teratocarcinoma; cancer; ss.
XX
OS Homo sapiens.
XX

```

```

FH Key Location/Qualifiers
FT CDS 115..1056
FT /tag=a
FT /product= BINDR
XX
XX MO9851795-A1.
XX
XX 19-NOV-1998.
XX
XX 13-MAY-1998; 98WO-US09877.
XX
XX 13-MAY-1997; 97US-0855518.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Corley NC, Hillman JL, Lal P;
XX
XX WPI: 1999-059701/05.
XX
XX P-PSDB: W87990.
XX
XX Double-stranded RNA-binding protein - used for treating or
XX preventing various cancers
XX
XX Claim 4; Fig 1A-E; 61pp; English.
XX
XX
XX The present sequence encodes a human double-stranded RNA-binding protein
XX (BINDR). The BINDR cDNA sequence was first identified in Incyte clone
XX 620438 from the paragon gillion cDNA library PGANN0701. BINDR can be used
XX in a pharmaceutical composition for treating or preventing cancer. The
XX BINDR polynucleotide can be used for the detection of polynucleotides
XX encoding BINDR. The cancers that can be treated using BINDR include
XX adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, and
XX teratocarcinoma, and particularly cancers of the adrenal gland, bladder,
XX bone, brain, breast, cervix, gall bladder, gastrointestinal tract, heart,
XX kidney, liver, lung, ovaries, pancreas, paragangliomas, parathyroid,
XX pituitary gland, prostate, salivary gland, spleen, stomach, thymus,
XX thyroid, testes, and uterus.
XX
XX Sequence 1608 BP; 481 A; 349 C; 344 G; 434 T; 0 other;

alignment_scores:
Quality: 219.50 Length: 471
Ratio: 1.081 Gaps: 17
Percent Similarity: 43.100 Percent Identity: 23.355

alignment_block:
US-09-316-048-11 x X03457 ..
Align seg 1/1 to: X03457 from: 1 to: 1608

54 GlyLysMetArgProProValIySAspAlaProAlaArgAlaLeuAr 70
111 ..... 111
39 GGGACGAGGAGCCGGGAGAGCAGTCGCTCCGACCGCTCTTCNC 88
70 gThr.....LeuGlnSerGluProLeuProGluArgLeuGluValA 84
88 ..... 88
84 snGlyArgGluAla.GluGluGluAsnLeu..... 93
139 GAGGCCCGCCGCTGGAGCGGAGCAGTGGACCTTCAGTTTGGGAA 188
94 .....AsnLySerGluIleSerGlnValPheGluI 104
189 GATGATTAACAGCTTAACGAGGAGAAACCGATTCAGTATTACAGCAAT 238
104 lAlaLeuLyS...ArgAsnLeuProValIaSnPheGluValAlaArgGlu 119
239 ACGGCATGAAGACAAAGACATCCCAAGTT...TATCAATGTGAAGATCT 285
120 SerGlyProProHisMetLySAsnPheValThrArgValSerValGlyI 136
::: ..... 111

```



CC multiple sclerosis, Grave's disease or myasthenia gravis.  
 XX Sequence 1904 BP: 576 A; 401 C; 385 G; 540 T; 2 other:

alignment\_scores:  
 Quality: 212.00 Length: 499  
 Ratio: 1.044 Gaps: 16  
 Percent Similarity: 40.681 Percent Identity: 22.244

alignment\_block:  
 US-09-316-048-11 x X24067 ..

Align seg 1/1 to: X24067 from: 1 to: 1904

```

46 GlyGlyGlnGlnPheAsnGlyLysGlyLysMetArgProProValIshI 62
|||||
17 GGTGGCGCGCTGANNMCTCCCGGACACGACCGCGGAGCATGCA 66
62 sAspAlaProAlaArgAlaLeuArgThrLeuGlnSerGluProLeuPro. 78
67 CCGTCGTCGACACGCTTCACTGAGCTTGGAACAGGAACTAATCTCCCTC 116
78 .....
117 CATCTCCAACTGCTTGACAAAGCTCGCTCTCTCCACGCGCTGAGCG 166
79 .....GluArgLeuGluValAsnGlyAr 86
167 TCCCTTCTCCGCTGTCGACGAGACGACGCGCGCGAGCGCCCGCGCT 216
86 gGluAlaGluGluGluAsnLeu..... 93
217 GGAGCGCGGAGGACAGTGGACCTTCAGTTGGGGAAGATGAACAGCTA 266
94 .....AsnLysSerGluLeuSerGlnValPheGluLeuLeuLys... 107
267 ACCGAGGGAACACCGATTCAGATTAACAGAAATAGCGCATGAAGAC 316
108 ArgAsnLeuProValAsnPheGluValAlaArgGluSerGlyProProH 124
|||||
317 AAGAACATCCCGCTT...TATGAATGTGAAGATCTGATGTCMAATACA 363
124 smetLysAsnPheValThrArgValSerValGlyGluPheValGlyGlu 141
|||||
364 CCGTCCGACCTTTCACCTTCAGAGTAACGCTGGTGCATTAACCTGCAC 413
141 llyGluGlyLysSerLysLysLysLysLysLysAlaAlaArgAlaVal 157
|||||
414 GTGAGGTACAGTAAGAAAGTGGCGAAACATAGAGCTGCAGAGCTGCC 463
158 LeuGluGlnLeuArgLeuProProLeuProAlaValGluArgVally 174
464 ATA..... 466
174 sProArgIleLysLysSerGlnProThrCysLysThrAlaProAsp. 190
|||||
467 ...AACATTTGAAGCCCAATGCAAGTATTGCTTTGCGAGTTCTCGACC 512
191 .....TyrgLysGlnGlyMetAsnProIle 198
513 CCTTAATGCGCTGACCTTCACAGCAACCAAGAACACGCTTAATCTATT 562
199 SerArgLeuAlaGlnIleGlnGlnAlaLysGlyLysGluProGluTy 215
|||||
215 rmetLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMet 232
613 TACCTTTTCCCGAGGAGGAGCTGCTCATAGAGAGATATACACAA 662
232 lvalLysValGlyLysIshIshIshThrAlaGluGlyValGlyThrAsn 248
|||||

```

```

663 TTTCAGAGCTAGATCATTTATGANAAGTGAAGAGGCGCATCAAAAAAG 712
249 ValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPheLy 265
|||||
713 CAAGCCAAAAGAAATGCTGCGAATAATTCTT..... 745
265 sValProGlnAlaGlnProAlaLysProAlaLeuLysSerGluGluYer 282
745 ..... 745
282 hrProValLysLysProGlyAspGlyArgLysValThrPhePheGluPro 298
746 .....GCCAAATTTAGTAAATATT 763
299 SerProGlyAspGluAsnGlyThrSerAsnLysAspGluGluPheArgme 315
|||||
764 TCTCCA.....GAGAACCCACATTTCT..... 784
315 rProTyrlSerHisGlnGlnLeuProAlaGlyIleLeuPrometValP 332
784 ..... 784
332 rGluValAlaGlnAlaValAlaGlyValSerGlnGly.....HisHis 345
785 .....TTACCAATGTAGTAGACATTTTATGATGTACTTGCGCATTTCC 829
346 ThrLysAspPheThrArgAlaAlaProAsnProAlaLysAlaThrValTh 362
830 TTGAGCAAT..... 838
362 rAlaMetIleAlaArgGluLeuLeuTyrglyThrSerProThrAlaG 379
839 .....TCTCCTGGGAAA 851
379 Lu...ThrIleLeuLysSerAsnIleSerSerGlyHisValProHisGly 394
852 AGATCAACCTTACTGAAAAGAACGCTCCTTAGT.....ATT 886
395 ProArgThrArgProSerGluGlnLeuTyrlTyrlLeuSerArgAlaGln 411
887 CCAAAATACGATTAACATCCAGCTGCTTAGTGAATTTGCCAAGAACAAAG 936
411 yPheGlnValAlaGlyTyrlLysAspPheProLys.....AsnAsnLysA 425
937 TTTTATATATAATATTGATATATAGATGACTGAGCCCAATGACAT 986
987 ATCAATGT.....CTTGCTGAACGTCTCCACACGCCCCCATCACAGCTGT 1030
425 snGluCysValSerLeuIleAsnCysSerSerGlnProProLeuValSer 441
442 HisGlyIleGlyLysAspValAlaGluSerCysHisAspMetAlaAlaLeu 458
1031 CATGGCTCCGATCTCTGTCGCAATGACAAAGATGATGACGTCACAA 1080
458 nIleLeuLysLeu.....Serg 465
1081 TCGTTTGCAGTATTAAAGATAATAGCAGAAAGAAAGTAATCTGAGACA 1130
465 luleuAspGlnGlnSerThrGluMetProArgThrGlyAsnGlyPro 480
1131 ACTTAAAAAATCTTTCAGTAGACATAAAAGATCCCGCTGGCCCC 1177

```









17 GGTGGGGGGCTGGANNCTCCCGGGGACGACGACCCGGGAGCATGCA 66  
150 sAspAlaAlaAlaLysAlaLeuArgIleLeuGlnAsnGluProLeuPro. 166  
67 CCGTCGTGCACGAGCTTCACTAGGCTGGAAACAGGAACCTAATCTCTC 116  
166 ..... 166  
117 CATCTCCAACTTGCTTGACAAAGCTCGCTTCCTCCAGCCGCTGAGCCG 166  
167 .....GluArgLeuGluValAsnGlyAr 174  
167 TCCCTTCTCCCATGTCAGAGACAGCAGCCGCGGAGCCCGCGCT 216  
174 gGluSerGluGluGluAsnLeu..... 181  
217 GGAGCGCGGACAGTGGAGCTTCAGTTGGGGAGATGATTAACGCTA 266  
182 .....AsnLysSerGluIleSerGlnValPheGluIleAlaLeuLys... 195  
267 ACCCAGGGAACACCGGATTTCAGATTTACAGCAATACGCGCATGAAGACC 316  
196 ArgAsnLeuProValAsnPheGluValAlaArgGluSerGlyProProh1 212  
317 AAGAACATCCAGTT...TATGAATGTGAAGATCTGATGTGCAATACCA 363  
212 shetLysAsnPheValIThrLysValSerValGlyGluPheValGlyGlu 229  
364 CGTGCCCATCTTCACCTTCAGATACCGTGTGACATTAACCTGCACAG 413  
229 IyGluLysSerLysLysIleSerLysLysAsnAlaAlaIleVal 245  
414 GTCAAGGTACAACTAGAAAGCTGCGCAACATAGACTGCAAGGCTGCC 463  
246 LeuGluGluLeuLysLysLeuProProLeuProAlaValGluArgVally 262  
464 ATTAACATTTTGAA.....GCCAA 483  
262 sProArgIleLysLysLysThrLysProIleValLysProGlnThrSerP 279  
484 TGCAGTATTTGCTTTCAGATTCCTGACCCCTTAATGCTGACCCCTTCCA 533  
279 roGluTyrGlyGlnGlyIleAsnProIleSerArgLeuAlaGlnIleGln 295  
534 AGCAACCAAGAACAGCTTAATCCTATTGTTGTTACAGAAATGGCT 583  
296 GlnAlaLysLysGluLysGluProGluTyrThrLeuLeuThrGluArgG1 312  
584 ATTCATCATGCTGAGACTTCTCTGAATATACCTTCCAGAGGAGGAG 633  
312 yLeuProArgArgArgGluPheValMetGlnValLysValGlyAsnHisT 329  
634 ACCTGCTCAAGAGAGAAATATCTACAATTTGCAGGCTAGACTATTTA 683  
329 hrAlaGluGlyThrGlyThrAsnLysLysValAlaLysArgAsnAlaAla 345  
684 TGGAAACTGGAAAGGGGCGATCAAAAGCAAGCAAAAGGATGCTGCT 733  
346 GluAsnMetLeuGluIleLeuGlyPheLysValProGlnArgGlnProth 362  
734 GAGAAATTTCTT..... 745  
362 rLysProAlaLeuLysSerGluGluLysThrProIleLysLysProGlyA 379  
745 ..... 745  
379 spGlyArgLysValThrPheGluProGlySerGlyAspGluAsnGly 395  
746 .....GCCAAATTTAGTAATTTCTCA.....GAGAACCCAC 778  
396 ThrSerAsnLysGluAspGluPheArgMetProTyrLeuSerHisGlnG1 412  
779 AATTCT..... 784

412 nLeuProAlaGlyIleLeuProMetValProGluValAlaGlnAlaValG 429  
785 .....TTAAACAAATGTAAGTGA 800  
429 IyValSerGlnGly.....HisHisThrLysAspPheThrArgAla 442  
801 GACATTTCTTAGATGTACTTGCAATTCCTTGAGGAAT..... 838  
443 AlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArgGluLe 459  
838 ..... 838  
459 uLeuTyrGlyGlyThrSerProThrAlaGlu...ThrIleLeuLysAsnA 475  
839 .....TTCCTCTGTGAAAGATCAACTACTGAAAGAA 872  
475 snIleSerSerGlyHisValProHisGlyProLeuThrArgProSerGlu 491  
873 GCCTCCTTAGT.....ATTCCAAT.....ACAGATTACATCCAG 907  
492 GlnLeuAspTyrLeuSerArgValGlnGlyPheGlnValGluTyrLysAs 508  
908 CTGCTTAGTGAATTCGCCAAGAACAGGTTTAATATTAACATATTGGA 957  
508 pPheProLysAsnAsnLysAsn...GluPheValSerLeuIleAsnCys 524  
958 TTTAGTGTAGTACGAGCGCCATGAGCAATATCAATCTTCTGTAACCT 1007  
524 erSerGlnProProLeuIleSerHisGlyIleGlyLysAspValGluSer 540  
1008 CCACGAGCCCATCAACAGATCTGTGATGCTCGGTATCTCTGTCGCAAT 1057  
541 CysHisAspMetAlaAlaLeuAsnIleLeuLysLeu..... 553  
1058 GCACAAAGTATGACGCTCACATCTTGCAGTATTTAAGATAATAGC 1107  
554 .....SerGluLeuAspGlnGlnSerThrGluMetP 564  
1108 AGAAGAAAGTAATCTGAGGACACTTAATAAATCTTCAGTAGACACATA 1157  
564 roArgThrGlyAsnGlyPro 570  
1158 AAAAGTTCCTCTGCGCCC 1177  
seq\_name: /cogn2\_6/ptodata/2/ina/6\_COMB.seq:us-09-234-613-95  
seq\_documentation\_block:  
; Sequence 95, Application US/09234613  
; Patent No. 6132973  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Bandman, Olga  
; APPLICANT: Shah, Purvi  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Yue, Henry  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES  
; NUMBER OF SEQUENCES: 98  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/234,613
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/933,750
FILING DATE: September 23, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE NUMBER: PF-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 95:
SEQUENCE CHARACTERISTICS:
LENGTH: 1904 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: THPLA2S08
CLONE: 2734573
US-09-234-613-95

```

```

alignment_scores:
Quality: 218.00      Length: 490
Ratio: 1.033         Gaps: 15
Percent Similarity: 43.061  Percent Identity: 22.041

```

alignment block:  
US-09-316-048-6 x US-09-234-613-95 ..

Align seg 1/1 to: US-09-234-613-95 from: 1 to: 1904

```

134 GYGLYGLNGLNPhasnGLYsGLYsThrArgLInAlaLalysH1 150
|||||
17 GGTGGGGGGCGCTGGANNCTCCCGGGGACGACGCCGGAGCATGCA 66
150 SASPALAIALALysALALeuArgIleuGlnAsnGluProLeuPro. 166
| : : : : : : : : : : : : : : : : : : : : : : : : :
67 CGTGTGCGACACGCTTCACTGAGGCTTGGACAGAACTAATCTCCCTC 116
166 ..... 166
117 CATCTCCACTGCTGACAAGCTGCTCTCCAGCCGCTGAGCCG 166
167 .....GluArgLeuGluValAsnGlyAr 174
167 TCCCTTCTGCCCATGTCGCCAGACGAGCAGCGCGCCGCCGCGCT 216
174 gGluSerGluGluGluAsnLeu..... 181
217 GGAGGGGAGACAGTGGACCTTCACTTGGGAAGATATAACACTA 266
182 .....AsnLysSerGluIleSerGlnValPheGluIleAlaLeuLys... 195
267 AGCCAGGGAACAACGCGATTACGATTAACGCAATGACGATGAAGACC 316
196 ArgAsnLeuProValAsnPheGluValAlaArgLInAsnGlyProProH1 212
317 AAGAACAATCCAGTT...TATGAATGGAAGAAGATCTATGCAATAACA 363
212 smELtYsAsnPhenValThrLysValSerValGlyGluPheValGlyLug 229
| : : : : : : : : : : : : : : : : : : : : : : : : :
364 CGTGCCCACTTTCACCTTCAGAGTAACGCTTGGTACATAACCTGCACAG 413
229 YGLUGLYLysSerLysLysIleSerLysAsnAlaIleAlaVal 245
|||||||
414 GTGAAGGTACAGTAAGAGCTGGCAACATAGAGCTGCAGAGGCTGCC 463

```

```

246 LeuGluGluLeuLysLysLeuProProLeuProAlaValGluArgVally 262
| : : : : : : : : : : : : : : : : : : : : : : : : :
464 ATAAACATTTTGA.....GCCAA 483
262 sPronArgIleLysLysLysThrLysProIleValLysProGlnThrSerp 279
| : : : : : : : : : : : : : : : : : : : : : : : : :
484 TGAAGATTATTTGCTTTCAGATTCCTGACCCCTTAAGCCGACCTTCCA 533
279 roGluYrGlyGlnGlyIleAsnProIleSerArgLeuAlaGlnIleGln 295
| : : : : : : : : : : : : : : : : : : : : : : : : :
534 AGCAACCAAGAAACAGCCTTAATCCATTGCTTCATACAGAAATGGCT 583
296 GlnAlaLysLysGluLysGluProGluYrThrLeuLeuThrGluArgL 312
584 ATTCATCATGCTGGAGACTTCCTCAATATACCTTCCACAGAGGAGG 633
312 yLeuProArgArgArgLupPheValMetGlnValGlyAsnHisT 329
| : : : : : : : : : : : : : : : : : : : : : : : : :
634 ACCTGCTCATTAAGAGAAATATCTACAATTTGCGAGCTAGACTATTTA 683
329 hralagLugLYThrGLYThrAsnLysLysValAlaLysArgAsnAla 345
684 TGGAAACTGGAAAGGGGCGCATCAAAAAGCAAGCAAAAGGAATGCTGCT 733
346 GluAsnMetLeuGluIleuGlyPheLysValProGlnArgGlnProth 362
734 GAGAAATTTCTT..... 745
362 rLysProAlaLeuLysSerGluGluLysThrProIleLysLysProGluY 379
745 ..... 745
379 spGlyArgLysValThrPhePheGluProGlySerGlyAspGluAsnGly 395
746 .....GCCAATTTTAGTAATATTTCCTCA.....GAGAACCCAC 778
396 ThrSerAsnLysGluAspGluPheArgMetProYrLeuSerHisGlnG 412
779 ATTTCT..... 784
412 nLeuProAlaGlyIleLeuProMetValProGluValAlaGlnAlaValG 429
785 .....TTACAAATGCTAGTAG 800
429 LyValSerGlnGly.....HisHisThrLysAspPheThrArgAla 442
801 GACATTTCTTAGAGTAGTACTTGGCATTCCTTGAGGAAT..... 838
443 AlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArgGluLe 459
838 ..... 838
459 uLeuYrGlyGlyThrSerProThrAlaGlu...ThrIleLeuLysAsnA 475
839 .....TCTCCTGCTGAAGAAAGATCACTACTGAAAAGAA 872
475 snIleSerSerGlyHisValProHisGlyProLeuThrArgProSerGlu 491
873 GCGCTCTTAGT.....ATTCCAAAT.....ACAGATTACATCCAG 907
492 GlnLeuAspYrLysSerArgValGlnGlyPheGlnValGluYrLysAs 508
908 CTGCTTAGTGAAATTTGCCAAGAAAGAGTTTATATATACATATTGGA 957
508 pPheProLysAsnAsnLysAsn...GluPheValSerLeuIleAsnCyss 524
958 TATAGATGAACCTGAGCCCAATGAGCAATATCAATGTCCTCAACCTGT 1007
524 erserGlnProProLeuIleSerHisGlyIleGlyLysAspValGluSer 540
1008 CCACAGCCCATCACAAGTCTGTGATGCTCCGGTATCTCTGTGGCAAT 1057
541 CyHisAspMetAlaAlaLeuAsnIleLeuLysLeuLeu..... 553

```

```

1058 GCAAGAGTATGACGCTGACATGCTTTGCAGATATTTAAAGATAATAGC 1107
554 ..... ||||| ||| |||::: |||
1108 AGAAGAAAGTAATATCGAGAGCAACCTTAAAAAATCTTCACTAGACATA 1157
564 roArghrGlyAaNglyPro 570
::: |||||
1158 AAAAGTTCCTCGCGCC 1177

seq_name: /cgn2_6/plodata/2/lna/5B_COMB.seq:US-08-855-518-2

seq_documentation_block:
Sequence 2, Application US/08855518
Patent No. 5858675
GENERAL INFORMATION:
APPLICANT: Billman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Neil C. Corley
TITLE OF INVENTION: DOUBLE-STRANDED RNA-BINDING
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/855,518
FILING DATE: Filed Herewith
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0239 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1608 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: PGANNOT01
CLONE: 620438
US-08-855-518-2

alignment_scores:
Quality: 211.50 Length: 404
Ratio: 1.143 Gaps: 13
Percent Similarity: 45.792 Percent Identity: 23.267

Alignment_block:
US-09-316-048-6 x US-08-855-518-2 ..
Align seg 1/1 to: US-08-855-518-2 from: 1 to: 1608

183 kVsErgIuIleSergIuValPheGluIleAlaLeuIys..ArgasnLe 198
||||: |||::: ||| :|||::: |||

```

211	AAACACCGCATTCACGGTATTACACGAATACGGCATGAAAGCCAAAGAAAT	260
198	uproValAsnpheGluValAlaArgGluSerGlyProProHisMetLysA	215
261	CCCACTT...TATGATGGAAGAAGATCTGATGGCAAAATACAGCTGCCCA	3070
215	snpheValThrLysValSerValGlyGluUpheValGlyGluGlyGlyGly	231
308	CTTTCACCTTCAGAACTAACCGTTGGAGCATTAACCGCACAGGTGAAGGT	357
232	LysSerLysLysLysSerLysLysAsnAlaAlaLeuAlaValLeuGlu	248
358	ACAAGTAAAGAAAGCTGCCGAACATAGAGCTGCAGAGCGTCCCTAAACAT	407
248	uLeuLysLysLeuProProLeuProAlaValGluArgValLysProArgI	265
408	TTTGGAA...GGCAATGCAAGTA	427
265	LeLysLysLysThrLysProIleValLysProGlnThrSerProGluTyr	281
428	TTTGCTTTCAGATTCCTGACCCCTTAATCCCTGACCCCTTCCCAAGCAACA	477
282	GlyGlnGlyLysAsnProIleSerArgLeuAlaGlnAlaGlnGlnAlaLys	298
478	AAGAACACAGCTTAATCCCTATGTGTCATTACAGAAATGGCTATTGCATCA	527
298	sLysGluLysGluProGluTyrThrLeuLeuThrGluArgGlyLeuProA	315
528	TGGCTGGAGACTTCCTGAATATACCTTTCCGAGGAGGAGCACTGCTC	577
315	rgArGArGluUpheValMetGlnValLysValGlyAsnHisThrAlaGlu	331
578	ATTAAGAGGAATATACCTACAAATTCGAGGCTGAGTCAATTTATGAAACT	627
332	GlyThrGlyLysAsnLysLysValAlaLysArgAsnAlaAlaGluAsnMe	348
628	GGAAGGGGGCATTCAAAAGCAAGCAAAAGAAATGCTGCTGAAGAAAT	677
348	LeuGluIleLeuGlyPheLysValProGlnArgGlnProThrLysProA	365
678	TCTT...681	
365	LeuLeuLysSerGluGluLysThrProIleLysLysProGluLysArgLys	381
682	...GCC	684
382	LysValThrPhePheGluProGlySerGlyAspGluAsnGlyThrSerAs	398
685	AAATTAGTAATATTCTCCA...GAAACACACTTTCT	720
398	nLysGluAspGluPheArgMetProTyrLeuSerHisGlnLeuProA	415
720	...720	
415	LeuGlyIleLeuProMetValProGluValAlaGlnAlaValGlyLysSer	431
721	...TTACAAATGTAGAGCAATCTT	744
432	GlnGly...HisIsthrLysAspPheThrArgAlaAlaProAs	445
745	TTAGCATGTACTGGCATTCCTTGAAGAAAT...744	
445	nProAlaLysAlaThrValThrAlaMetIleAlaArgGluLeuLeuTyrG	462
774	...774	
462	LysGlyThrSerProThrAlaGlu...ThrIleLeuLysAsnAsnIleSer	477
775	...TCTCCTGGTGAAGAAGATCAACTCTGAAAGAAAGACCTCTT	816
478	SerGlyHisValProHisGlyProLeuThrArgProSerGluGlnLeuAs	494
817	AGT...ATTCCAAAT...ACAAGTTATCATCAGCTGGCTTGA	851



```

1892 AATCAGAGAAATCATCCATATTCACAGAGAAAGAAATCAGAGAAC 1941
272 eValysProGlnThr.....SerProGluTyrGlnGly 285
1942 TGAGAGATCCAGAGACCCCTTCAGCAGACATCTTTCTGGGA 1991
285 LeuSerProIleSerArgLeuAlaGlnIleGlnGlnAlaLysGluLys 301
1992 AGAGCCCCCGACACACATCTTGATGTATGCAC...AATGGGGGAC 2038
302 GluProGluTyrThrLeuLeuThrGluArgGlyLeuProArgArgGly 318
2039 TCCTGGCAATTCCTCTCTCTCCCAAGAGGCCCTGCATGACCCAA 2088
318 uphValMetGlnValLysValGlnHisThrAlaGlnGly...Thrg 334
2089 GTTCCAAATACTGTGTTCAGTGGAGCCCAACTTCCCGAGTGTAGTG 2138
334 LysThrAsnLysValAlaLysArgAsnAlaAlaGlnAsnMetLeuGlu 350
2139 CTCGCCAGCAAGAAAGTGCGAAAGCAGATGGCCGACAGAGAACCATGAG 2188
351 IleLeu.....GlyPheLysValProGlnArgGlnPro... 361
2189 GCCCTCATGGGAGGAGCAGCAATCATGCTTGATTAACACGCTGA 2238
362 .....ThrLysProAlaLeuLysSerGluGlnGlyThr 373
2239 AGGTATGATCTCAGAGTCACTGATTAATCTTGATTCATGATGATCCCAACA 2288
373 rolleLysLysProGlyAspGlyArgLysValThrPhePheGluProGly 389
2289 AGGTCAAGAAAGATTGGCAGACTGCTGATACCTGAACACCAACCTGTG 2338
390 SerGly..... 391
2339 GGTGGCTTTTGAGTACGCCCTCCATGGCTTGTCTGATTAACAA 2388
392 .....AspGluAsnGlyThrSerAsnLysGluAspGluPheArgMetP 406
2389 GTTGTGACAGTCCGACCTCTCTCAGAG.....C 2420
406 rotLysLeuSerHisGlnLeuProAlaGly...IleLeuProMetVal 421
2421 CCAAGTTCCTTACCAAGCAAAAGTTGGGGTGGCTGCCCAAGCCGTC 2470
422 .....ProGluValAlaGlnAla..... 427
2471 TGGCAGACAGCAAGAGCAAGCAGACAGCAGATGCGGCTCT 2520
428 .....ValGlyValSerGlnGlnHisThrLysAspPheThr 441
2521 CCGTGTCTTGATTTGGGAGAGCAAGAGCAGAGCATGGTTTCACAG 2570
441 rgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArg 457
2571 AGGTAAACCCCA.....GTGACAGGGCCACACTCTCAGAGA 2605
458 GluLeuLeuTyrGlyLysThrSerProThrAlaGluThrIleLeuLysAs 474
2606 ACATATCTCTCTCTCAAGAGTCCCAAGACAG..... 2641
474 nasnIleSerSerGlyHisValProHisGlyProLeuThrArgPro.... 489
2642 .....CCAAAGACACTCTCTCTCA 2660
490 .....SerGluGlnLeuAspTyrLeuSerArgValGlnGly 501
2661 CTGGCAGACACTTCATGACCAAGATGACATGCTGTAGCCAC..... 2701
502 PheGlnValGluTyrLysAspPheProLysAsnAsnLysAsnGluPheVal 518
2702 .....CGGTGCTTCA 2712

```

```

518 IserLeuIleAsnCySerSerGlnProProLeuIleSerHisGlyIleG 535
2713 CACTGTGACTAAC...AGCTTCAGCCCTCTCTGCTGGCCGCAAGATT. 2758
535 LysAspValGluSerCyHisAspMetAlaAlaLeuAsnIleLeuLys 551
2759 .....CTGGCCGCCATC...ATTATGAAA 2779
552 LeuLeuSerGluLeuAspGlnGlnSerThrGluMetProArgThrGlyAs 568
2780 AAGACTCTGAG...GACATGGGTGTGCTGTCAGCTTGGGAGACAGGAA 2826
568 nglyPrometSerValCysGlyArgCys 577
2827 T.....CGCTGT 2833

seq_name: /cgn2-6/plodata/2/lna/5A_COMB.seq:US-08-457-459-1
seq_documentation_block:
? Sequence 1, Application US/08457459
? Patent No. 5677428
? GENERAL INFORMATION:
? APPLICANT: Nishikura, Kazuko
? TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
? TITLE OF INVENTION: Thereof
? NUMBER OF SEQUENCES: 39
? CORRESPONDENCE ADDRESS:
? STREET: Howson and Howson
? CITY: Spring House
? STATE: Pennsylvania
? COUNTRY: USA
? ZIP: 19477
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/457,459
? FILING DATE:
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/197,794
? FILING DATE: 17-FEB-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/280,443
? FILING DATE: 25-JUL-1994
? ATTORNEY/AGENT INFORMATION:
? NAME: Bak, Mary E.
? REGISTRATION NUMBER: 31,215
? REFERENCE/DOCKET NUMBER: WST49CUSA
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 215-540-9206
? TELEFAX: 215-540-5818
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 6671 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: unknown
? MOLECULE TYPE: CDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 155...3832
? US-08-457-459-1

alignment_scores:
? Quality: 157.00 Length: 626
? Ratio: 0.553 Gaps: 31
? Percent Similarity: 45.367 Percent Identity: 22.364

```

alignment block:

US-09-316-048-6 x US-08-457-459-1 ..

Align seq 1/1 to: US-08-457-459-1 from: 1 to: 6671

```

33 MetSerIleProSerThrThrSerSerLeuProSerGluAsnAlaGlyAr 49
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1241 ATGCAATTCAGAGAAATACGACAGCTGTTCTT...GAACCGCTCCAGC 1287
49 gProIle.....GlnAsnSerAlaLeuProSerAlaSerIleT 62
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1288 TGCATTCCTGAGACCAAGAAAGACGAGAGTCTCCTCACTGTAATATAC 1337
62 hrSerThrSerAlaAlaGluSerIleThrProThrValGluLeuAsn 78
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1338 CCACATCAAAATGCTCAATATACATGTTACACGAGAAAAGTGAGAGAT 1387
79 AlaLeuGlyMetLeuGlyLysLysProMetThrLysProValAspPr 95
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1388 .....GGCGAGAACCTGCTCATMAAG.....TT 1410
95 cTyrSerArgMetGlnSerThrTyrAsnTyrAsnMetArgGlyGlyAlaT 112
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1411 AGAAACAGCGCAGAGAGGACGA..... 1432
112 yTrpProAlaGlyTrpPheThrProPheProVal.....ProProLeu 125
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1433 .....CCAGAACCGACAGACAGCTGAACACCACTGTT 1462
126 LeuTyrGlnValGluLeuSerValGlyGlyGlnIlePhe...AsnGlyLy 141
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1463 CATTCACATGCGCCCTCAAAAGCAGAGGTATGTGACTTTGAATGGCGCA 1512
141 sGlyLysThr.....ArgGlnAla 148
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1513 GTGGCGCAGCATGACATCCAGATGACTGTAATAGTATCCGCGCAGCAC 1562
148 IalysHisAspAlaAlaAlaLysAlaLeuArgIleLeuGlnAsnGluPro 164
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1563 CAGGTGAGTTTCGAGCCATCATGAGAGATGCCCTCTCTTACAGTCATGCC 1612
165 LeuProGluArgLeuGluValAsnGlyArgGluSerGluGluAsnIle 181
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1613 TTGGCA.....CGTGTTCACCCCTACAGAAACT 1641
181 uAsn.....LysSerGluIleSerGlnValPheGluIleAla 194
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1642 GACGAGGTGCGCAGCTGAAGAACCCCATCAGCGGCTTTAGAAATATGCC 1691
194 euLysArgAsnLeuProValAsnPheGluValAlaArgGluSerGlyPro 210
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1692 AGTTTCGCTAGCAAACTGTGATTCACATGATAGACAGACAGTGCACCA 1741
211 ProHisMetLysAsnPheValThrLysValSerVal...GlyGluPheVa 226
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1742 CCCCATGACCTCGATTAAATTCAGAGTTGTCAATCATGCGCCAGAGATT 1791
226 IGlYgluGlyGluGlyLysSerLysLysIleSerLysLysAsnAlaAla 243
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1792 TCCCCCAGCTGACCTGCAAGAAAGTGGCCAGACAGATGACGCTTA 1841
243 le.....AlaValLeuGluGluLeuLysLysLeuProProLeu 255
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1842 TGAAGACCATACAAATTCCTAGAGGAAAGCCAAAGCCAGACAGTGCAG 1891
256 ProIlaValGluArgValLysProArgIleLysLysLysThrLysProI 272
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1892 AAATCAGAGAAATCATCCCACTATTCCACAGAGAAAGATCAGAGAAC 1941
272 eValLysProGluThr.....SerProGluTyrGlyGlnGlyI 285
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1942 TGCAGAGTCCAGACCCCTTCAGCCACATCTTCTTTCTTGCGGA 1991

```

```

285 leAsnProIleSerArgPheAlaGlnIleGlnAlaLysLysGluLys 301
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1992 AGAGCCCGCTACACACTGCTTGAAGTATGCAC...AAATTGGGAGAAC 2038
302 GluProGluTyrThrLeuLeuThrGluArgGlyLeuProArgArgArg 318
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2039 TCTCGGCAATTCCTGCTCTCTGCAAGAGAGCCCTGCCATGAAGCCAA 2088
318 uPheValMetGlnValLysValGlyAsnHisThrAlaGluGly...Thr 334
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2089 GTTCCAAATACCTGTGTGCGAGTGGAGCCCAAACTTCCCGAGTGAAGT 2138
334 lyThrAsnLysLysValAlaLysArgAsnAlaAlaGluMetLeuGlu 350
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2139 CTCCGAGCAAGAAAGTGGCAAGATGCGCGCAGAGAGAACCATGAG 2188
351 IleLeu.....GlyPheLysValProGluArgGlnPro.. 361
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2189 GCCCTGATGGGAGGAGCGACCACTCATGGGCTTCTGATTAACCACTGA 2238
362 .....ThrLysProAlaLeuLysSerGluGluLysThrP 373
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2239 AGGTATGATCTCAGAGTCACCTGTGATTAATCGATGATGCCACAA 2288
373 roIleLysLysProGluValAspGlyArgLysValThrPhePheGluPro 389
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2289 AGCTCAGAGAAATTTGGCGAGCTCGAGATACCTGAACCAACCACTGTG 2338
390 SerGly..... 391
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2339 GGTGGCTTTGGAGTACGCGCCGCTCCCATGCTTGTGCTGATTAACA 2388
392 .....AspGluAsnGlyThrSerAsnLysGluAspGluPheArgMet 406
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2389 GTTGTGACAGCAGTCCGACCTCTCTACAGAG.....C 2420
406 roTyrLeuSerHisGlnIleLeuProAlaGly...IleLeuProMetVal 421
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2421 CCAGTGTGTTTACCAACCAAAAGTTGGGGGTGCTGTTCCAGCCGTC 2470
422 .....ProGluValAlaGlnAla..... 427
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2471 TGGCGACAGACAGACAGCAAGGACAGAGACAGATGCGCTCT 2520
428 .....ValGlyValSerGlnGlyHisHisThrLysAspPheThr 441
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2521 CCGTGTCTTGATTGGGAGAGAGAGAGAGCATGGGTTTCCACAG 2570
441 rGAlaAlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArg 457
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2571 AGGTAAACCCA.....GTGACAGGGGCCAGTCTCGAAGAA 2605
458 GluLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLys 474
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2606 ACTATGCTCTCTCTCTCAAGGTCCAGAGACAGAG..... 2641
474 nasIleSerSerGlyHisValProHisGlyProLeuThrArgPro... 489
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2642 .....CCAAAGACACTCTCTCTCA 2660
490 .....SerGluGlnLeuAspTyrLeuSerArgValGlnGly 501
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2661 CTGGCAGACACTTCATGACAGATAGCATAGCTTAGGCCAC..... 2701
502 PheGlnValGluTyrLysAspPheProLysAsnAsnLysAsnGluPheVa 518
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2702 .....CGGTGCTTCA 2712
518 lSerIleLeuAsnGlySerSerGlnProProLeuIleSerHisGlyIle 535
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2713 CACTCTGACTAAC...AGCTTCACAGCCCTCTGCTGCGCCAGAGATT. 2758

```

535 lylsyaspyalgluSerCysHisAspMetAlaAlaLeuAsnIleLeuLys 551  
:::|||||::: |||:::|  
2759 .....CTGCGCCGCATC...ATTATGAAA 2779  
552 LeuLeuSerCyluLeuAspGlnGlnSerThrGluMetProArgThrGlyAs 568  
||||| |||::: |||  
2780 AAGACTCTTAG...GACATGGGTGTCTGTACGTGGTGGACAGGAA 2826  
568 nglyProMetSerValCysglYArgCys 577  
| |||||  
2827 T.....CGCTGT 2833  
seq.name: /cgn2\_6/ptodata/2/lna/5A\_COMB.seq:US-08-555-678-1  
seq\_documentation\_block:  
: Sequence 1, Application US/08555678  
: Patent No. 5763174  
: GENERAL INFORMATION:  
: APPLICANT: Nishikura, Kazuko  
: TITLE OF INVENTION: RNA Editing Enzyme and Methods  
: TITLE OF INVENTION: of use thereof  
: NUMBER OF SEQUENCES: 67  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Howson and Howson  
: STREET: Spring House Corporate Cntr, P.O. Box 457  
: CITY: Spring House  
: STATE: Pennsylvania  
: COUNTRY: USA  
: ZIP: 19477  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/555,678  
: FILING DATE:  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/197,794  
: FILING DATE: 17-FEB-1994  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/280,443  
: FILING DATE: 25-JUL-1994  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/457,459  
: FILING DATE: 01-JUN-1995  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Bak, Mary E.  
: REGISTRATION NUMBER: 31,215  
: REFERENCE/DOCKET NUMBER: WST49DUSA  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 215-540-9206  
: TELEFAX: 215-540-5818  
: INFORMATION FOR SEQ ID NO: 1:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 6671 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: double  
: TOPOLOGY: unknown  
: MOLECULE TYPE: CDNA  
: FEATURE:  
: NAME/KEY: CDS  
: LOCATION: 155...3832  
: US-08-555-678-1  
alignment\_scores:  
Quality: 157.00 length: 626  
Ratio: 0.553 Gaps: 31  
Percent Similarity: 45.367 Percent Identity: 22.364  
alignment\_block:

US-09-316-048-6 x US-08-555-678-1 ..  
Align seg 1/1 to: US-08-555-678-1 from: 1 to: 6671  
33 MetSerIleProSerThrThrSerSerLeuProSerGluAsnIleGlyArg 49  
|||||::: |||::: |||::: |||  
1241 ATGCATCAATCAAGAAATACGACAGCTGTC...GAAACCGCTCCAGC 1287  
49 gProIle.....GlnAsnSerAlaLeuProSerAlaSerIle 62  
|||::: |||::: |||::: |||  
1288 TGCATATCCCTGAGACCAAGAAAGAACCGACAGATCTTCACCTGTAATATAC 1337  
62 hrSerThrSerAlaAlaAlaGluSerIleThrProThrValGluLeuAsn 78  
::: |||::: |||::: |||  
1338 CCACATCAATGCGCTCAATACATGCTACACAGAAAAAGTGAGAAAT 1387  
79 AlaLeuCysMetLysLeuGlyLysLysProMetLysLysProValAsp 95  
|||||::: |||::: |||  
1388 .....GGCAGGAACCTGTCTATAAG.....TT 1410  
95 oYrSerArgMetGlnSerThrTyraSerValAsnMetArgGlyValAT 112  
::: |||::: |||  
1411 AGAAACAGGCAAGAGCCAGA..... 1432  
112 yrProProArgTyrrPheTyrrProPheProVal.....ProProLeu 125  
||| |||::: |||  
1433 .....CCAGAACACAGACAGACAGTGAACACCACTGT 1462  
126 LeuTyrrGlnValGluLeuSerValGlyGlyGlnGlnPhe...AsnGlyLys 141  
|||::: |||::: |||  
1463 CATTCACATGGGCCCTCAAAAGCAGCGTATGTGACTTGAATGGCCCA 1512  
141 sGlyLysThr.....ArgGlnAla 148  
||| |||  
1513 GTGGCCACAGATGACATCCACATGACTTGAATGATCCGCGCAC 1562  
148 lAlsyHisAspAlaAlaAlaLysAlaLeuArgIleLeuGlnAsnGluPro 164  
::: |||::: |||  
1563 CAGGTGAGTTGCGCATCATGAGAGTCCCTCTCATGACTATGGC 1612  
165 LeuProGluArgLeuGlnValAsnGlyArgGluSerGluGluAsnLe 181  
||||| ||| |||  
1613 TTGCCA.....CGGTGTCCACCCCTACAGAAACT 1641  
181 uAsn.....LysSerGluIleSerGlnValAlaPheGluIleAla 194  
|||::: |||::: |||  
1642 GACAGAGTCCGACCTGAAGAACCCCATCAGCGGCTGTGAATATGCC 1691  
194 euLysArgAsnLeuProValAsnPheGluValAlaArgGluSerGlyPro 210  
::: |||::: |||  
1692 AGTTCCTACTCAAACTGTGAGTTCACATGATAGAGCAGAGTGACCA 1741  
211 ProHisMetLysAsnPheValThrLysValSerVal...GlyLysPheVal 226  
|||||::: |||::: |||  
1742 CCCCATGAACTGATTAATTCAGGTTGTCATCAATGCCGAGAGT 1791  
226 lGlyGlnGlyGluLysSerLysLysIleSerLysLysAsnAlaAla 243  
::: |||::: |||  
1792 TCCCCCAGCTGAAGCTGGAAGCAAGAAAGTGCCCAAGAGATGCAGCTA 1841  
243 le.....AlaValLeuGluGluLeuLysLysLeuProProLeu 255  
::: |||::: |||  
1842 TGAAGCCATGACAAATTCCTAGAGGAAGCAAGCAAGACAGAGTGA 1891  
256 ProAlaValGluArgValLysProArgIleLysLysLysThrLysPro 272  
::: |||  
1892 AATTCAGAAAGATCAATCCACTATCCACAGAGAAAGAAATCAGAGAAC 1941  
272 eValLysProGlnThr.....SerProGluTyrrGlyGlnGly 285  
|||||::: |||  
1942 TGCAGAGTCCAGACCCCACTTCACAGCATCTCTTTTCGGGA 1991  
285 leAsnProIleSerArgLeuValGlnIleGlnAlaLysLysGluLys 301



```

1992 AGAGCCCCGTCACACACTCTTGATGTATGCAC...AAATTGGGAAC 2038
302 G1uPrGluThrLeuThrGluArgGlyLeuProArgArgGly 318
2039 TCCTGGCAATTCGTCCTCTCCAAAGAGCGCTGCCCATGAACCCAA 2088
318 upheValMetGlnValLysValGlnAsnHisThrAlaGluGly...Thrg 334
2089 GTTCCATACACTGTGTTCAGTGGAGGCCAAACTTTCGCCAGTGTAGTG 2138
334 LThrAsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGlu 350
2139 CTCGCCAGCAAGAAAGTGGCAAGACGATGGCCGAGAGAGCAACCATGAG 2188
351 IleLeu.....GlyPheLysValProGlnArgGlnPro... 361
2189 GCCCTGCATGGGAGCGCACCACTCCATGGCTTGTATTAACGACCTGA 2238
362 .....ThryProAlaLeuLysSerGluGluLysThrP 373
2239 AGGATGATCTCAGAGTCACTTGATTACTTGATGATCATGATGCCAACA 2288
373 roLleLysLysProGlyAspGlyArgLysValThrPhePheGluProGly 389
2289 AGGTCAAGAAATTTGGCGAGCTGTGATACCTGAACACCAACCTGTG 2338
390 SerGly..... 391
2339 GGTGGCCCTTTTGGAGTACGCCCGCTCCATGGCTTGTCTGTAATTCAA 2388
392 .....AspGluAsnGlyThrSerAsnLysGluAspGluPheArgMetP 406
2389 GTTGTGTGACACGACGCGACCTCTCTCACAG.....C 2420
406 roTyrlLeuSerHisGlnGlnLeuProAlaGly...IleLeuProMetVal 421
2421 CCAAGTTCGTTTACCAAGCAAAAGTTGGGGTGTCTGTTCCACGCGTC 2470
422 .....ProGluValAlaGlnAla..... 427
2471 TGGCCACACAGCAAGAAAGCAAGCAACGACGACGATGCGGCTCT 2520
428 .....ValGlyValSerGlnGlnLysHisThrLysAspPheThrA 441
2521 CCGTGTCTTGTATGGGAGAAAGCAAGAGCAAGCATGGTTTCACAG 2570
441 rgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetLleAlaArg 457
2571 AGGTAAACCCCA.....GTGACAGGGGCCAGCTCAGAGA 2605
458 GluLeuLeuTyrgLysGlyThrSerProThrAlaGluThrIleLeuLysAs 474
2606 ACTATGCTCTCTCTCAAGTCCCGCAGAGCAG..... 2641
474 nasIleSerSerGlyHisValProHisGlyProLeuThrArgPro.... 489
2642 .....CCAAAGACACTCCCTCTCA 2660
490 .....SerGluGlnLeuAspTyrlLeuSerArgValGlnGly 501
2661 CTGCGACACACTTTCATGACCATGACGATGCTGAGCCAC..... 2701
502 PheGlnValGluTyrlLysAspPheProLysAsnLysAsnGluPheVal 518
2702 .....CGTGTCTTCAA 2712
518 LserLeuIleAsnLysSerSerGlnProProLeuIleSerHisGlyIleG 535
2713 CACTCTGACTAAC...AGCTTCCAGCCCTCTGCTGGCGGCAAGATT. 2758
535 LysAspValGluSerCysHisAspMetAlaAlaLeuAsnIleLeuLys 551

```

```

2759 .....CTGGCCGCCATC...ATTATGAAA 2779
552 LeuLeuSerGluLeuAspGlnGlnSerThrGluMetProArgThrGlyAs 568
2780 AAACACTGTAG...GACATGGTGTCTGTCTGACGCTTGGGAAAGGAA 2826
568 nglyPrometSerValCysGlyArgCys 577
2827 T.....CGCTGT 2833

seq_name: /cgn2_6/pdata/2/lna/PCTUS_COMB.seq:PCT-US95-02275-1

seq_documentation_block:
: sequence 1, Application PC/TUS9502275
: GENERAL INFORMATION:
: APPLICANT: Wistar Institute of Anatomy & Biology
: TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
: NUMBER OF SEQUENCES: 39
: CORRESPONDENCE ADDRESS:
: ADDRESSER: Howson and Howson
: STREET: Spring House Corporate Cntr, P.O. Box 457
: CITY: Spring House
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19477
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/02275
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/280,443
: FILING DATE: 25-JUL-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/197,794
: FILING DATE: 17-FEB-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Bak, Mary E.
: REGISTRATION NUMBER: 31,215
: REFERENCE/DOCKET NUMBER: WST49BPCT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 215-540-9206
: TELEFAX: 215-540-5818
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6671 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: unknown
: MOLECULE TYPE: CDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 155..3832
: PCT-US95-02275-1

alignment_scores:
Quality: 157.00 Length: 626
Ratio: 0.553 Gaps: 31
Percent Similarity: 45.367 Percent Identity: 22.364

alignment_block:
US-09-316-048-6 x PCT-US95-02275-1 ..
Align seq 1/1 to: PCT-US95-02275-1 from: 1 to: 6671
33 MetSerIleProSerThrThrSerSerLeuProSerGluAsnAlaGlyAr 49

```

```

1241 ATGCAATCAAGAGAAATACAGACAGTGTCT...GAAACCGCTCCAGC 1287
49 gpProlle.....GlnAsnSerAlaLeuProSerAlaSerIleT 62
1288 TGCATTCCTGAGACCAAGAAAGCAGAGTTCCTACCTGTAATATAC 1337
62 hrSerhrSerAlaAlaIaIaLuserIleThrProThrValGluLeuAsn 78
1338 CCACATCAAAATGCTCAATATACATGTATACACAGAAAAAGTGGAAT 1387
79 AlaLeuCysMetLysLeuGlyLysLysProMetLysLysProValAsp 95
1388 .....GGGCGAGAACCTGCTCAATAG.....TT 1410
95 oTySerArgMetGlnSerThrTyrAsnTyrAsnMetArgGlyAlaIaT 112
1411 AGAAACAGGCGAAGAGCCAGA..... 1432
112 yrProProArgTyrPheTyrProPheProVal.....ProProlou 125
1433 .....CCAGAACCCAGACAGACTGAAACCACTGTT 1462
126 LeuTyrGlnValGluLeuSerValGlyGlyGlnGlnPhe...AsnGlyLys 141
1463 CATTCAGATGGCCCTCAAAAGCAGGGTATGTGACTTGAAATGGCCA 1512
141 sgLysLysThr.....ArgGlnAla 148
1513 GTGGGCCACAGATACATCCAGATGACTGTAATGATTCGGCGACAC 1562
148 lAlaHisAspAlaAlaAlaLysAlaLeuArgLLeuGlnAsnGluPro 164
1563 CAGGTAGATTGAGCCATCATGAGATGCCCTTCCTACACTCATGCGC 1612
165 LeuProGluArgLeuGluValAsnGlyArgGluSerGluGluAsnLe 181
1613 TTGCCA.....CGCTGTTCACCCCTACCAAGAAACT 1641
181 uAn.....LysSerGluLysSerGlnValPheGluIleAla 194
1642 GACAGAGTCCAGCTGAGAGAACCCCATACGCGGCTGTTAGAATATGCC 1691
194 euLysArgAsnLeuProValAsnPheGluValAlaArgLysSerGlyPro 210
1692 AGTTCCTAGTCAACCTGTGAGTTCACATGATAGACAGAGTGGCCA 1741
211 ProHisMetLysAsnPheValThrLysValSerVal...GlyLubPheVal 226
1742 CCCCATGAACTCGATTTAATTCAGGTTTCATCATGCGCGAGATT 1791
226 lGlyGluGlyGluGlyLysSerLysLysLysSerLysAsnAlaIaT 243
1792 TCCCCAGCTGAGAGCTGAGAGCAAGAAAGTGGCCAGACAGATGACGTA 1841
243 le.....AlaValLeuGluGluLeuLysLysLeuProProlou 255
1842 TGAAGACCATGACAAATTCGTGAGAGAAAGCCAAAGCCAAAGACAGTGA 1891
256 ProAlaValGluArgValLysProArgLLeLysLysLysThrLysPro 272
1892 AAATCGAAGAAATCATCCACTATTCACAGAGAAAGAAATCAGAGAGAC 1941
272 eValLysProGlnThr.....SerProGluTyrGlyGlyGlyT 285
1942 TGCAGAGTCCAGACCCACCCCTTCAGCCACATCCTTTCTTGGGA 1991
285 leAsnProLysSerArgLeuAlaGlnIleGlnGlnAlaLysLysGlyLys 301
1992 AGAGCCCGCTACACACATGCTGATGATGAC...AAATTGGGGAGC 2038
302 GluProGluTyrThrLeuLeuThrGluArgGlyLeuProArgArgArg 318
2039 TCCTGGCAATTCGTCCTCTCTCCAAAGAGGCCCTGCCATGAAACCCA 2088

```

```

318 uPheValMetGlnValLysValGlyAsnHisThrAlaGluGly...ThrG 324
2089 GTTCCAAATACTGTGTTCAGATGGAGCCCAACTTTCCCAAGTGTAGTG 2138
334 lYThrAsnLysValAlaLysArgAsnAlaIaLglnAsnMetLeuGlu 350
2139 CTCGCCAGCAAGAAAGTGGCAAGACAGATGGCCGACAGAGAAAGCATGAG 2188
351 lLeu.....GlyPheLysValProGlnArgPro... 361
2189 GCCCTCATGGGAGGCGACCAACTCCATGGCTTGTGATTAACGACGCTGA 2238
362 .....ThrLysProAlaLeuLysSerGluLysThrP 373
2239 AGGTATGATCTCAGACTCATTGATTAAGTTCATGATGATGATGCCAACA 2288
373 rolLeuLysProGlyAspGlyArgLysValThrPhePheGluProGly 389
2289 AGGTGAGAAAGATTTGGGAGCTGTGAGATACCTGAACACCAACCTGTG 2338
390 SerGly..... 391
2339 GGTGGCTTTTGGAGTACGCCCTCCATGGCTTGTGCTGAATTCAA 2388
392 .....AspGluAsnGlyThrSerAsnLysGluAspGluPheArgMetP 406
2389 GTTGTGACACACTCCGACCTCTCACAG.....C 2420
406 roTyrLeuSerHisGlnGlnLeuProAlaGly...lLeuProMetVal 421
2421 CCAGTTTCGTTTACCAAGCAAAAGTTGGGGTGTGCTGTCCACAGCGTC 2470
422 .....ProGluValAlaGlnAla..... 427
2471 TGGCACACAGCAAGAGCAAGCAAGAGAGAGACAGATCGCGCTCT 2520
428 .....ValGlyValSerGlnGlnHisThrLysAspPheThr 441
2521 CCGTCTTGATTTGGGAGACAGAGCAAGAGCAAGCAAGCAAGTTCACAG 2570
441 rGlnAlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArg 457
2571 AGTTAACCCA.....GTGACAGGCGCCAGTCTCAGAAAG 2605
458 GluLeuLeuTyrGlyLysThrSerProThrAlaGluThrIleLeuLys 474
2606 ACTATGCTCTCTCTCAAGGTCCAGAGACAG..... 2641
474 nasIleSerSerGlyHisValProHisGlyProLeuThrArgPro.... 489
2642 .....CCAAAGACACTCCCTCTCA 2660
490 .....SerGluGlnLeuAspTyrLeuSerArgValGlnGly 501
2661 CTGGCAGCACTTCATGACAGATAGCATGCTGCTGACCCAC..... 2701
502 PheGlnValGluTyrLysAspPheProLysAsnAsnLysAsnGluPheVal 518
2702 .....CGGTGCTTCAA 2712
518 lSerLeuIleAsnCysSerSerGlnProProlouLysSerHisGlyIleG 535
2713 CACTGTGACTAAC...AGCTTCAGGCCCTCTGTCGCGCGCAATTT. 2758
535 lLysAspValGluSerCysHisAspMetAlaAlaLeuAsnIleLeuLys 551
2759 .....CTGGCCGCAATC...ATTATGAAA 2779
552 LeuLeuSerGluLeuAspGlnGlnSerThrGluMetProArgThrGlnLys 568
2780 AAGACTCTGAG...GACATGGGTGTGTCTGTCAGCTTGAGGAACAGGAA 2826

```



```
|||||
2125 ... AGTGGACAGGAATGACATCACACCCACCCAGCAAGAG... 2166
402 lnpheargmetprotyrleuSerHisGlnleuProAlaGlyIleu 418
2167 .....CTC 2169
419 PrometValProGlu.....ValAlaGlnAlaVala 428
2170 CCACCAAGCCGACAGAGAAACAAAGCCTTGCCACACCTACCTGC 2219
428 lclValSerGlnGlyHisHisThrLysAspPheThrArgAlaAlaPro 445
2220 AAGACCTTCACATCGAAGCCAAACACACGCCACTCTCTCCCTAACG 2269
445 snProAlaLysAlaThrValThrAlaMetIleAlaArgGluLeuLeuTyr 461
2270 ACCCAAGCTCCACACCATGGTGGTGAATAAAMACCATG..... 2313
462 GlyGlyThrSerProThrAlaGluThrIleLeuLysAsnAsnIleSer 478
2314 .....AGCCTTCTTC 2324
478 rGlyHisValProHisGlyProLeuThrArgProSer 490
2325 AGGCTTAGTGCCAGCTGCCACCCAAACGCCCTGCC 2361

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:us-08-770-379-20

seq_documentation_block:
: Sequence 20, Application US/08770379
: Patent No. 5849564
: GENERAL INFORMATION:
: APPLICANT: Chang, Yuan
: APPLICANT: Bohenzky, Roy A.
: APPLICANT: Russo, James J.
: APPLICANT: Edelman, Isidore S.
: APPLICANT: Moore, Patrick S.
: TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
: TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Cooper & Dunham LLP
: STREET: 1185 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/770,379
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: White, John P.
: REGISTRATION NUMBER: 28,678
: REFERENCE/DOCKET NUMBER: 52342
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 278-0400
: TELEFAX: (212) 391-0525
: INFORMATION FOR SEQ ID NO: 20:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 32207 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-770-379-20
```

```
alignment_scores:
  Quality: 132.50      Length: 496
  Ratio: 0.616        Gaps: 24
  Percent Similarity: 43.347  Percent Identity: 20.161

alignment_block:
US-09-316-048-6 x US-08-770-379-20
Align seg 1/1 to: US-08-770-379-20 from: 1 to: 32207

36 ProSerThrThrSerSer.....LeuProSerGluAsnAla..... 47
|||||
4882 CCAGATAGTACTCTCTCTCTGAGCTTGCCCAAGATCCGAGCATTCG 4931
48 .....GlyArgProIleG 52
4932 CCTCCCTGGCCGCCGACGACGAGAGAGCGTGCCACTGGGAACCCGTGC 4981
52 lAsnSerAlaLeuProSerAlaSerIleThrSerThrSerAlaAla 68
|||||
4982 GCCACATTCACACCCCGCA.....AGAGATACAGATCTTTCGCG 5022
69 GluSerIleThrProThrValGluLeuAsnAlaLeuCysMetLysLeuG 85
5023 TCAAGCGCGCTGGCTTCATATACCTTCGCAATTA..... 5058
85 yLysLysProMetTyrLysProValAspProTyrSerArgMetGlnSer 101
5059 ..CCAGATCCCGACGACCGCAACCGATACGATCGATGCTGTAACGGT 5107
102 ThrTyrAsnTyrAsnMetArgGly.....Al 111
5108 .....AACAAATACCTTGAAACACGCTACATCCCTCGGATCGCAAGC 5151
111 aTyrProProArgTyrPheTyrProPheProValProProLeuLeuTyrG 128
5152 GGACCGCGTACGACCATCTATGATGTCCTCGTAAGCG..... 5191
128 lValGluLeuSerValGlyGlyGlnGlnPheAsnGlyLysGlyLysThr 144
5192 .....CAGGTGATGG 5203
145 ArgGln.....AlaAlaLys 149
5204 CCTGAATATATGACACACAGCAACACCCCTGCGCTACGACATCGCG 5253
149 sHisAspAlaAlaAlaLysAlaLeuArgIleLeuGlnAsnGlu..... 163
5254 AAGAGAGCGGAGTACAGTTCGCCGACTCTTCAGAGCCAGGTGAGCG 5303
164 .....ProLeuProGluArgLeuGluValAsn 172
5304 CCACATGCTGTACATCTTCCAGAGGTCCGAGCGCTCTCCCTCCCAT 5353
173 GlyArgGluSerGluGluGluAsnLeuAsnLysSerGluIleSerGlnVal 189
5354 GCGCCTCAGAGAGTTCACGACATG.....TCGAAAT 5388
189 lPheGluIleAlaLeuLysArgAsnLeuProValAsnPheGluValAla 206
5389 CCACGAGGTGCGC.....GCCAGCGCGCCACTCACAATCACCCCAATA 5432
206 rGluSerGlyProProHisMetLysAsnPheValThrLysValSerVal 222
5433 AACCGACCGAAGCCCTCAGTCTCCCGAGCGTGAATCAATACAGAA 5482
223 GlyGluPheValGlyGluGlyLysSerLysLysIleSerLysLys 239
5483 GCGCAA.....CGCGACGACGACCGCAAGATGTCGCGGA 5517
239 sAsnAlaAlaIleAlaValLeuGluLeuLysLysLeuProProLeu 256
|||||
```

```

5518 CAACATCCCTAGTGTATTCCTCCGTCGCTA.....CCGACCC 5555
256 roalaValaIgluArgValaIgluProaIgluLysLysThrLysProIle 272
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5556 CG.....AAACCCAGAGAGCGCTAGATCAGCTACGCCAC 5590
273 VallysProgluThrSerProgluTyrGlyGlnGlyIleasnProIleSe 289
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5591 GCACGTGCGCCGTTATATCACCCCA...GCACATGCGCCGCTGCTATAC 5637
289 rArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProIuTyr 306
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5638 GCATCTG.....CCAGCAGCGCAGTAACGAGCCCAAGAGGG 5675
306 hrLeuLeuThrGluArgGlyLeuProaIgluArgGluPheValMetGln 322
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5676 TTCCTCAAAGCAACGTGA..... 5695
323 ValLysValaIgluAsnIsthAlaGluGlyThrGlyThrAsnLysLysVa 339
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5696 .....ACTCT 5700
339 lAlaLysArgAsnAlaIgluAsnMetLeuGluIleLeuGlyPheLysV 356
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5701 CGTCTGCGCGCGCGCG.....GTCA 5723
356 alProgluArgGluProThrLysProaIleuLysSerGlnGluLysThr 372
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5724 TTGACCCAGAGAGCGGCTCTGCGCACCAGATCCGATATGAGAGAGCG 5773
373 ProIleLysLysPro.....GlyAspGlyArgLys..... 382
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5774 GCGCTCCAGCGCGCGGAGTGAAGGAGGAGCGCGCGCGCTCCGACAC 5823
383 .....ValThrPhePheGluProGlySerGlyLysPheLysang 395
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5824 GCAACCGCTCACTTAACCTTCTCTCCACCTACCGCACCCACCTCCCG 5873
395 lYThrSerAsnLysGluAspGluPheArgMetProTyrLeuSerHisGln 411
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5874 CACTGGAGCCCTAGAAACCAAAACAACCTCCCA..... 5908
412 GlnLeuProaIgluIleLeuProMetValProgluValaIgluAlaVa 428
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5909 TTCACGCGCCCGACACGCCATAGACATTAGCCA.....CCACAGACACC 5952
428 lGlyValSerGlnGlyHisIsthLysAspPheThrArgAlaIleProA 445
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5953 TCCCATGTCCACTCCTCCTCAGCGAGACCAAGCCCGCGCAGAAA 6002
445 snProaIleAlaThrValThrAlaMetIleAlaArgGluLeuLeuTyr 461
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6003 AGCGGCGCGACCCGCTCTTCGAGTAATG..... 6031
462 GlyLysThrSerProThrAlaGluThrIleLeuLys 473
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6032 GCGCCCAAGCAACGTCGGGAGAGCAAGATCAAG 6067
seq_name: /cgn2_6/ptodata/2/ina/5a_COMB.seq:US-08-143-219-1
seq_documentation_block:
; Sequence 1, Application US/08143219
; Patent No. 5670330
; GENERAL INFORMATION:
; APPLICANT: Sonenberg, Nahum
; APPLICANT: Katze, Michael G.
; APPLICANT: Roy, Sophie
; APPLICANT: Koromilas, Antonis E.
; APPLICANT: Barber, Glen N.
; TITLE OF INVENTION: TUMOR-CELL ASSAY METHOD AND KIT
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon

```

```

; STREET: 611 West Sixth Street
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM compatible
; OPERATING SYSTEM: PC-DOS (Version 5.0)
; SOFTWARE: Wordperfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/143,219
; FILING DATE: October 25, 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION DATA: including application
; APPLICATION NUMBER: 08/141,244
; FILING DATE: October 22, 1993
; APPLICATION NUMBER: 07/953,681
; FILING DATE: September 29, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Douglas E. Olson
; REGISTRATION NUMBER: 22,798
; REFERENCE/DOCKET NUMBER: 204/139
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2628 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: HUMAN PKR GENE, FIGURE 5
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 187..1836
; US-08-143-219-1
;
; alignment_scores:
; Quality: 130.00 Length: 293
; Ratio: 0.949 Gaps: 12
; Percent Similarity: 46.758 Percent Identity: 23.891
;
; alignment_block:
; US-09-316-048-6 x US-08-143-219-1 ..
;
; Align seg 1/1 to: US-08-143-219-1 from: 1 to: 2628
;
195 LysArgAsnLeuProValAspPheGluValAlaArgLysSerGlyProPr 211
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
244 AAGCAGGAGTAGTACTTAATATCAAGAACTGCTTAATTCAGGACCTCC 293
227 lYgluGlyGluGlyLysSerLysLysIleSerLysLysAsnAlaIle 243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
344 CAGAAGGTCAAGTAGATCAAGAAAGAAAGCAAAATAATGCGCAGCAAA 393
244 AlaValLeuGluGluLeuLysLysLeuProProLeuProAlaValGluAr 260
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
394 TTAGCTGTGACATCTTAATAG..... 417
260 gValLysProaIgluLysLysLysThrLysProIleValLysProgluT 277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

418 .....GAAAGAGCAGTACTCTTATTATTGACAA 454
277 hrserPro.....GluTyrGlyGlnGlyIleAsnProIleSer 289
    |||:|||||:|||||:|||||:|||||:
455 CGAATCTTCAGAGAGATTTCATGGGGAATTACATAGCCTTATCAAT 504
290 ArgLeuAlaGlnIleGlnAlaLysGluLysGluProGluTyrTrh 306
    |||:|||||:|||||:|||||:|||||:
505 AGAATTGCCAG.....AGAAAAGACTAACTGTAATTATGA 542
306 rLeuLeuThrGluArgGlyLeuProArgArgGluPheValMetGln 323
    |||:|||||:|||||:|||||:|||||:
543 ACAGTGTGCATCG...GGGTGCATGGCCGAGAGATTTCATTATAAT 589
323 alLysValGly...AsnHisThrAlaGluGlyTrhGlyTrhAsnLys 338
    |||:|||||:|||||:|||||:|||||:
590 GCAAAATGGCAGAGAAATATAGTATTGCTACAGTCTTACTAAACAG 639
339 ValAlaLysArgAsnAlaAlaGlu...AsnMetLeuGluIleLeuGly 354
    |||:|||||:|||||:|||||:|||||:
640 GAAGCAAAACAATTGGCCGCTAAACTTCATCTTCAGATATTA..... 684
354 elysValProGlnArgGlnProThrLysProAlaLeuLysSerGlnGlu 371
    |||:|||||:|||||:|||||:
685 .....TCGAAAGAA. 693
371 ystrhrProIleLysLysProGlyAspGlyArgLysValThrPhePheGlu 387
693 ..... 693
388 ProGlySerGlyAspGluAsnGlyTrhSerAsnLysGluAspLysPhe 404
    |||:|||||:|||||:|||||:
694 .....ACCTCAGTGAATCTGCAC..... 711
404 gMetProTyrLeuSerHisGlnIleLeuProAlaGlyIleLeuProMet 421
    |||:|||||:|||||:|||||:
712 .....TACCTGTCC..... 720
421 alProGluValAlaGlnAlaValAlaGlyValSerGlnGlyHisThrLys 437
721 .....TCTGGT 726
438 AspPheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAla 454
    |||:|||||:|||||:|||||:|||||:
727 TCTTTGCTACGACTGCTGAGTCCCAAGCAACTCTTTAGTACAGCAC 776
454 tIleAlaArgGluLeuLeuTyrGly.....ThrSerProThrAlaG 469
    |||:|||||:|||||:|||||:|||||:
777 ACACGCTTCTGAAATCATCTGAGAGTGCATCTTCAGCAGATACATCAG 826
469 luthrIleLeuLysAsnAsnIleSerSer 478
    |||:|||||:|||||:|||||:|||||:
827 AGATTAATCTACAGTGCAGGTTTAACAGT 855

```

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:us-08-258-261B-6

seq\_documentation\_block:

```

; Sequence 6, Application US/08258261B
; Patent No. 5639949
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/258, 261B
; FILING DATE: 08-JUN-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457, 205
; FILING DATE: 01-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; US-08-258-261B-6

```

```

alignment_scores:
    Quality: 128.00      Length: 656
    Ratio: 0.489        Gaps: 30
    Percent Similarity: 39.939    Percent Identity: 20.427

```

alignment\_block:

US-09-316-048-6 x US-08-258-261B-6 ..

Align seg 1/1 to: US-08-258-261B-6 from: 1 to: 28958

```

18 SerGlnIleLeuAsnLysAsnGlnSerLeuLeuSerGlnProLeu..... 32
    |||:|||||:|||||:|||||:|||||:
10574 ACATCGCTCTCGAGCTGCAGCTCCCTCTTATTATCGACCGCACCGGC 10623
33 .....MetSerIleProSerThrThrSerSerLeuProSerGluAsnA 47
    |||:|||||:|||||:|||||:|||||:
10624 ACCAGCGCTGCGCTCGAGCTGCAGCGCGCGCTGCTGATGCAACCT 10673
47 lacyArgProIleGlnAsnSerAlaLeuProSerAlaSerIleThrSer 63
    |||:|||||:|||||:|||||:|||||:
10674 CCGCAACCGCTCTCTCTCGAGCGCGCACCGCGCTCTCGACGAGG 10723
64 ThrSerAlaAlaGlnSerIleThrProThrValGluLeuAsnAlaLe 80
    |||:|||||:|||||:|||||:|||||:
10724 GGCATGCTTCTCCGTCGAGGTGACGCC..... 10753
80 ucysMetLysLeuGlyLysLysProMetLysProValAsp..... 94
    |||:|||||:|||||:|||||:|||||:
10754 ....ATCCCGTGTCTACGCTCGCCCTCCCGAGACCTGGAGCGGTCA 10799
95 .....ProTyrSer ArgMetGlnSerThrTyrAsnTyrAsnMetLar 108
    |||:|||||:|||||:|||||:|||||:
10800 GCTCATCCGCTGCTGCTGCGCTCCATTCGACGACAGAGA..... 10838
108 gGlyGlyAlaTyrProProArgTyrTrhPheTyrProPheProValPro 125
    |||:|||||:|||||:|||||:|||||:

```

10839 .....AGGCACCTCGC.....CCGCTGCTCTC 10863  
125 eufeytrglInValGluLeuSerValGlyGlnGlnPheasnGlyLys 141  
10864 TCCTG.....GGCGAGCTCTCTAC..... 10883  
142 GlyLysThrArgGlnAlaAlaLysHisAspAlaAlaAlaLysAlaLeu 158  
10884 .....CCGAGGCTCGCGCTGACGTGAAGACTCTTCGCGCCCTACG 10927  
158 glleleuGln.....AsnGluProLeuProGluArgLeu... 169  
10928 CTCGCCGAGAGGTCTCTCTCCCTACCTACCTCCAGAGAGCGTTC 10977  
170 .....GluValAsnGlyArg..... 174  
10978 TGCGCTGAGCTCTCCAGAGAGACGCTTCGAGCTCCGCTCCGAGGCC 11027  
175 .....GluSerGluGlnGluAsnLeuAsnLys 183  
11028 TGACCTGCGCGCCGACCAATCCCGCTGCTCGCGCGCGCTTCGCGC 11077  
183 sSerGluLysSerGlnValPheGluLeuAlaLeuLysArgAsnLeuPro 200  
11078 ACCGCGGTGGCTTCTCTTACAGGGCGCTCTCTCCGAGAGCAACCG 11127  
200 aL..... 200  
11128 TGCGCTGAAGGCATCGCTCTTCGCGACACCCATCTACCGGGACCGG 11177  
201 .....AsnPh 202  
11178 CTTTCGAGCTCGCCCTGCGACGTGCGCCACCGCGTGGCTCGACACCG 11227  
202 egluValAlaArgGluSerGly.Pro.....ProHisMetLysAsn 215  
11228 TCGAAGGCTCAGCTCGAGGCCCTCTGCTCTCCATCGCAGAGACCC 11277  
216 pheValThrLysValSerVal.GlyGluPheValGlyGlnGlyL 232  
11278 GTCCTCTCAGATCTCTCGGCGCGCGGTGAGACGAGACGAGAGGGC 11327  
232 ySerLysLysLysSerLysLysAsn..... 240  
11328 GCTCTCTTCATAGCCGACAGAGAGAGCGCTTCAGATGCGCCCTGGA 11377  
241 ..AlaAlaIleAlaValLeuGlnGluLeuLysLysLeuProProLeu 256  
11378 CTCGCCAGCGCAGCGCTCTCTCTCGCGCGACCCCATCTCTCGCGC 11427  
256 alaValAlaArgValLysProArgIleLysLysThrLysProIle 273  
11428 GATCTCCAGAGTGGCTCTCTGAGTGCATCGCGGTGAGACTCGAAGG 11477  
273 alLysProGlnThrSerProGluTyrGlyGlnGlyLysAsnProIleSer 289  
11478 CCTCTACGGAACCTCGCAACTCGGGCTTCGCTACGCGCGAGGTTC 11527  
290 ArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTyrTh 306  
11528 AGGGCTTC..... 11536  
306 rleuLeuThrGluArgGlyLeuProArgArgGluPheValMetGln 323  
11536 ..... 11536  
323 alLysValGlyAsnHisThrAlaGlnGlyThrGlyThrAsnLysVal 339  
11537 .....GCTCGCTCAACAAGCGCGAGCGAGCTCTTCGCGAAGCCA 11578  
340 AlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPheLysVa 356  
11579 AGCTCCCG.....AAGCG 11592

356 lProGluArgGlnProThrLysProAlaLeuLysSerGluGluLysThr. 372  
11593 GCCGAAAGATGCGC.....CCGGTTGCCCTCCACCTGCGCTGCT 11636  
373 .....ProIleLysLysProGluLysArg 381  
11637 CGACAGCGCCCTGATGACCTGGCTTTGAGAGACGACAGAGAGGAGCG 11686  
382 LysValThrPhePheGluProGlySerGlyAspGluAsnGlyThrSerAs 398  
11687 TCGCTTCG.....CCTCTGAGAGCGAGTCTGCTGCTGCT. 11725  
398 nLysGluAspGluPheArgMetProTyrLeuSerHisGlnGluPro 415  
11725 ..... 11725  
415 laGlyIleLeuProMetValPro.GluValAlaGlnAlaValGlyLys 431  
11726 .....CCGTGCTGACACACCTTCGCGGTGCTTCACCGCTC 11764  
431 rgIngly..... 433  
11765 CCAAGGTGAATCTCGTCTGATGCTGCTGCGCGAGCGAGGTGAC 11814  
434 .....HisThrLysAspPheThrArgAla 442  
11815 CCTCTGCTGCTGCTGACGCGCTGCGATGCGAGAGCGTCCGCGCGCA 11864  
443 AlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArgGlu 459  
11865 GTCGCGACCC..... 11876  
459 uLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysAsn 476  
11877 .....GGCAGCTTCCA...CCATGATGCGCTCTTCGCGTGAC 11913  
476 leSerSerGlnHisValProHis.....GlyProLeuThrArg 488  
11914 TGAGAGCAGCTCCAAAGCCCACTTCACCGCTGCGCGCGCGAGCGCT 11963  
489 ProSerGluGlnLeuAspTyrLeuSerArg.ValGlnGlyPheGlnVal 505  
11964 CTTCTCGGACAGAGCGCGCACGATCTCGCGCTGAGCGCCCGCTGCC 12013  
505 lutyLysAspPheProLysAsnAsnLysAsnGluPheValSerLeuIle 521  
12014 GCTACGCGGACCTCGCTGCGCTCGAAGC.....GCCCTCGAC 12051  
522 AsnCySerSerGlnPro.....ProLeuIleSerHisG 533  
12052 CAGGCGCTTCGCTCGCGCTGCTGCTGCTGCTGCTTCATCGATCACC 12101  
533 yIleGlyLysAspValGluSerCyHisAspMetAlaAlaLeuAsnIle 550  
12102 GCGAGGAGACCTGCTCGGAGCGCGCACGAGGCGACGCGCTGCACTCG 12151  
550 euLysLeu 552  
12152 CCCTCTTG 12159  
seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-456-837-6  
seq\_documentation\_block:  
; Sequence 6, Application US/08456837  
; Patent No. 5643774  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ilgon, James M.  
; APPLICANT: Beck, James Joseph  
; APPLICANT: Hill, Dwight Steven  
; APPLICANT: Ryals, John Andrew  
; APPLICANT: Gaffney, Thomas Deane

APPLICANT: Lam, Stephen Ting  
APPLICANT: Hammer, Phillip E.  
APPLICANT: Uknes, Scott Joseph  
TITLE OF INVENTION: Genes for the synthesis of  
TITLE OF INVENTION: antipathogenic substances  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESS: Ciba-Geigy Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/456,837  
FILING DATE: 01-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/457,205  
FILING DATE: 01-JUN-1995  
APPLICATION NUMBER: 08/258,261  
FILING DATE: 08-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8614  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 28958 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-456-837-6

alignment\_scores:  
Quality: 128.00 Length: 656  
Ratio: 0.489 Gaps: 30  
Percent Similarity: 39.939 Percent Identity: 20.427

alignment\_block:  
US-09-316-048-6 x US-08-456-837-6 ..  
Align seg 1/1 to: US-08-456-837-6 from: 1 to: 28958

18 serglnlleuasnlnysasnlnserlleuaserlnproleu..... 32  
10574 ACATCGCTCTCGAGCGTCGAGCTCTCTTTATTTCGACCGACCGGC 10623  
33 .....MetSerlleProSerThrThrSerSerleuProSerGluAsnA 47  
10624 ACCAGGCTCGAGCGCTCCGACGTCGACGCGCGGTACTGTATCGAACCT 10673  
47 laGlyArgProIleGlnAsnSerAlaLeuProSerAlaSerIleThrSer 63  
10674 CCGGCAACCGCTCTCTCTCGACGCGACGACGCGCTCTCGACGANG 10723  
64 ThrSerAlaAlaGluSerIleThrProThrValGluLeuAsnAlaLe 80  
10724 GGCATCGCTTCTCCGTCGAGTCAGCCGCC..... 10753  
80 uCysMetLysLeuGlyLysLysProMetTyrLysProValAsp..... 94

10754 ...ATCCCGTCACGCTCCGCCCTCCCGGAGACCTCGACCGCTACCC 10799  
95 .....ProTyrSer ArgMetGlnSerThrTyrAsnTyrAsnMetAr 108  
10800 GCTCGATCCCGTCGTCGTGCGCTCCATTCGACGAGAGA..... 10838  
108 gLylGlyAlaTyrProProArgTyrPheTyrProPheProValProProL 125  
10839 .....AGCCACCTCCG.....CCGCTGCTCTC 10863  
125 euLeuTyrGlnValGluLeuSerValGlyGlyGlnGlnPheAsnGlyLys 141  
10864 TCCCTG.....GGCGAGCTCTCTAC..... 10883  
142 GlyLysThrArgGlnAlaAlaLysHisAspAlaAlaAlaLysAlaLeuAr 158  
10884 .....CCGAGGCTCGCGCTCGACTGGAAGACTTCTTCCGGCCCTACG 10927  
158 gLleLeuGln.....AsnGluProLeuProGluArgLeu.... 169  
10928 CTCGCCGCAAGGTCTCCCTCCACCTACCTCCTCCAGGAGAGCGGCTTC 10977  
170 .....GluValAsnGlyArg..... 174  
10978 TGGCTCGACGCTCTCCAGGACGACGACTTCCGACGCTCGCCTCCGACGCC 11027  
175 .....GluSerGluGluGlnAsnLeuAsnLys 183  
11028 TCACCTCGCGCCGACCAATCCCGCTCGCGCGCGCGCTTCTTGGCG 11077  
183 sserGluIleSerGlnValPheGluIleAlaLeuLysArgAsnLeuProV 200  
11078 ACCGCGGTGGCTTCTCTTACAGGCGGCTCTCCCTCCAGAGCACCGG 11127  
200 al..... 200  
11128 TGGCTCGAGGCCATGCGCTTTCGCGACACCCATCTACCGGACCGG 11177  
201 .....AsnPh 202  
11178 CTTTTCGAGCTCGCCCTCGACGTCGCCACGCGCTCGGCTCGACACG 11227  
202 eGluValAlaArgGluSerGly.Pro.....ProHisMetLysAsn 215  
11228 TCGAAGAGCTCAACGCTCGAGGCGCTCTCGCTCTCCATTCGACGACG 11277  
216 PheValThrLysValSerVal.GlyGluPheValGlyGluGlyLys 232  
11278 GTCCTCTCCAGATCTCCGTCGCGCGCGGTGACGACGACGAGAGGCG 11327  
232 ySerLysLysIleSerLysAsn..... 240  
11328 GCTCTCTTTCATAGCCGACAGAGAGCGCGCTTCAGATGCGCCCTGGA 11377  
241 ...AlaAlaIleAlaValLeuGluGluLeuLysLysLysLeuProProLeuPr 256  
11378 CTCGCCACCCAGCGGCTCTCTCTCGCGGCGACCCCATCTCTCTCGCG 11427  
256 oAlaValGluArgValLysProArgIleLysLysLysThrLysProIleY 273  
11428 GATCTCCAGAGTGGCTCTCCGATGATGCCATCCCGGTGACCTCGAAG 11477  
273 alLysProGlnThrSerProGluTyrGlyGlnGlyLysProIleSer 289  
11478 CCTTACGGAACCTTCGCAACCTCGGCTTCCCTACGAGCCCGGACTTCC 11527  
290 ArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTyrTh 306  
11528 AGGCGCTCC..... 11536  
306 rLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMetGlnV 323



```

11536 ..... 11536
323 alysva1g1yasn1h1s1h1a1g1u1g1y1h1r1g1y1h1r1a1n1s1y1s1v1a1 339
      ::::: ::::: :::::
11537 .....GCTCGGTCTACAGAGCGGCGGAGAGCTCTTGGCCGAGACCA 11578
340 AlAlaYsArGAsn1A1a1a1g1u1a1s1n1m1e1u1g1l1e1u1g1y1p1h1e1y1s1v1a 356
      :::::
11579 AGCTCCCGG.....AAGCG 11592
356 lProGlnArGlnProThr1ySProAla1e1u1ySerg1u1g1u1y1h1r. 372
      :::::
11593 GCCGAAAAGAGATGCG.....CCCGGTTGGCCCTCCACCCCTGGCGTCT 11636
373 .....Pro1le1y1s1ySProG1y1Asp1y1a1rG 381
      :::::
11637 CGACAGCGCCCTGCATGCATGGCCTTTGAGAGACAGACAGAGAGGAGCG 11686
382 LysVal1ThrPhePheGluProG1ySerg1yAsp1y1a1sng1y1h1rSeras 398
      :::::
11687 TCGCTCTGC.....CCTTCTCGAGAGAGAGTCTCGTGCCT.. 11725
398 n1ySg1u1aSp1uPheArGmetPro1y1r1e1uSern1h1sng1n1e1uProA 415
11725 ..... 11725
415 laG1y1le1e1uPro1e1tValPro.1GuVal1a1a1a1a1y1a1g1y1a1s1e 431
      :::::
11726 .....CCGTGGTGCACACACCTTGCGGTGCGCTTCCACCGCTC 11764
431 rG1ng1y..... 433
      :::::
11765 CCAAGGTAATGCTCGTCTGCATGCTCTGGCGACGCGGAGAGTAC 11814
434 .....His1h1r1y1aSp1h1e1h1r1a1r1a1a 442
      :::::
11815 CCGTCTGCCTGGTGCAGAGCGCTCGCCATCGGAGCGACGCGCGGCA 11864
443 AlAProAsnProAla1y1a1h1r1a1h1a1m1e1t1le1a1r1g1u1e 459
      :::::
11865 GCTCCGACACCC..... 11876
459 u1e1u1y1r1g1y1h1rSergPro1h1a1a1u1h1r1le1e1u1y1a1s1n1 476
      :::::
11877 .....GGCAGCTTCCA...CCATGATGGCTCTTCCGGGTGCAC 11913
476 lSerSerg1y1h1s1v1a1ProH1s.....G1y1Pro1e1u1Th1rArG 488
      :::::
11914 TGGAGCGAGCTCCAAAGCCCATCTTACCGCCCTGCGCCGAGCGGCGT 11963
489 ProSerg1u1n1e1u1aSp1y1r1e1uSera1rG.Va1G1nG1y1PheG1nVa1G 505
      :::::
11964 CCGTCTCGGAGCGGCGGACAGATCTCGGCTCGAGCGCCCGCTCCGCC 12013
505 l1u1y1r1y1aSp1h1e1h1r1y1a1s1n1y1a1sng1u1Phe1a1Ser1le1u1 521
      :::::
12014 GCTACGCCGACCTCGCTGCCCTCGGAAGC.....GCCCTGAC 12051
522 Asn1ySerg1e1rG1nPro.....Pro1e1u1le1Ser1h1s1G1 533
      :::::
12052 CAGGCGCGCTTCCGCTCCGCGCTCGTGCCTTTCATTCATTCAGTCAACC 12101
533 y1l1eG1y1y1aSp1y1a1g1u1Ser1y1h1s1aSp1e1t1a1a1a1e1u1a1n1e1l 550
      :::::
12102 GGCAGGCGAGCTCGTCCGAGCGCCACAGAGCGCACCGCGCTCGCACTCG 12151
550 e1u1y1s1e1u 552
      :::::
12152 CCCTCTTG 12159
seq_name: /cgn2_6/ptodata/2/1na/5A.COMB.seq:US-08-457-342-6
seq_documentation_block:

```

```

; Sequence 6, Application US/08457342
; Patent No. 5662898
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,342
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOHETICAL: NO
; ANTI-SENSE: NO
; US-08-457-342-6

alignment_scores:
  Quality: 128.00      Length: 656
  Ratio: 0.489         Gaps: 30
  Percent Similarity: 39.939   Percent Identity: 20.427

alignment_block:
  US-09-316-048-6 x US-08-457-342-6 ..

Align seg 1/1 to: US-08-457-342-6 from: 1 to: 28958

18 serGln1leuAsn1yAsn1g1n1Ser1e1uSerg1nPro1e1u..... 32
   :::::
10574 ACATCGCTCTCGAGAGTGGAGAGCTCTTTATTCAGACGTCACCGCGC 10623
33 .....MetSer1leProSer1h1r1h1r1Ser1e1uProSer1u1a1s1nA 47
   :::::
10624 ACCAGGCTCGAGCGGCTCGAGCTCGAGCGGCGTACTGATCGAAGACT 10673

```

[illegible]

11478	CCCTACGCAACCTCGCAACCTCGGGCTTCCTACGGCCCCGAGTTCC	11527
290	ArgLeuAlaGlnIleGlnIleAlaLysLysGluLysGluProGluTyrThr	306
11528	AGGGCGCTCC.....	11536
306	rLeuLeuThrGluIatGlyLeuProArgArgArgGluPheValMetGlnv	323
11536	.....	11536
323	allLysValGlyAsnHisThrAlaGluGlyThrGlyThrAsnLysLysVal	339
11537	.....GCTCCGCTCTACAAACCGCGGCGACGACGCTTTGGCCGAAGCCA	11578
340	AlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPheLysVal	356
11579	AGCTCCCG.....AACGG	11592
356	lProGlnArgGlnProThrLysProAlaLeuLysSerGluLysThr	372
11593	GGCGAAAGGATGCCG.....CCGGATTGGCTTCACCCCTGCGTCTCT	11636
373	.....ProIleLysLysProGluLysPheLysArg	381
11637	CGACAGCGCCCTGCATGCACCTGCGCTTCAGAGACAGACAGAGGAGACGG	11686
382	LysValThrPhePheGluProGluSerGlyAsnGluLysThrSerAs	398
11687	TCGCTTGC.....CCTTCTCGTGGACGGAGATCTCCGCTGCCT..	11725
398	nLysGluAspGluPheArgMetProTyrLeuSerHisGlnIleuProA	415
11725	.....	11725
415	IaGlyIleLeuProMetValPro.GluValAlaGlnAlaValaGlyLys	431
11726	.....CCGTGGTGCACACACCTTGGCGGTGCCTCCACCGCTC	11764
431	rGlnGly.....	433
11765	CCAGGGGTAAATCTCCGCTCGATCGTCTGCGCGGACCGCCAGGTGAC	11814
434	.....HisHisThrLysAspPheThrArgAla	442
11815	CCCTCTCGCTCGATGCAACGCGCTCGCATCGAGACGATCGCGCGCGCA	11864
443	AlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArgGluLe	459
11865	GCCTCCGACACCC.....	11876
459	uLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysAsnAsnI	476
11877	.....GGCAGCTTCCCA..CAATGATGCCGCTCTTCGCGCTCCAC	11913
476	LeuSerGlyLysValProHis.....GlyProLeuThrArg	488
11914	TGGAGGAGACTCCAAAGCCCACTTACAGCGCTCGCGCCCGGAGGGCGT	11963
489	ProSerGluGlnLeuAspTyrLeuSerArg.ValGlnGlyPheGlnValG	505
11964	CCCTTCTCGGACAGCGGCGGACAGATCTCCGCTCGACGCGCCGCTCGCC	12013
505	IuTyrLysAspPheProLysAsnAsnLysAsnGluPheValSerLeuIle	521
12014	GCACGCCGACCTCGCTCGCTCCGAGC.....GCCCTCGAC	12051
522	AsnCysSerSerGlnPro.....ProLeuIleSerHisGlnI	533
12052	CAGGGCGCTTGGCGCTCGCGGCTCGTGCCTGCGGCCCTTCATGATGACAC	12101
533	YIleGlyLysAspValGluSerCysHisAspMetAlaIleuAsnIleL	550

12102 GGCAGCGACCTGTCGCCAGGCCACGAGCCACCGCGCTGCGACTCG 12151

550 eulysleu 552

|||

12152 CCCTCTTG 12159

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-457-646A-6

seq\_documentation\_block:

Sequence 6, Application US/08457646A

Patent No. 5679560

GENERAL INFORMATION:

APPLICANT: Schupp, Thomas

APPLICANT: Ligon, James M.

APPLICANT: Beck, James Joseph

APPLICANT: Hill, Dwight Steven

APPLICANT: Ryals, John Andrew

APPLICANT: Gaffney, Thomas Deane

APPLICANT: Lam, Stephen Ting

APPLICANT: Hammer, Phillip E.

APPLICANT: Uknes, Scott Joseph

TITLE OF INVENTION: Genes for the synthesis of

TITLE OF INVENTION: antipathogenic substances

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: Ciba-Geigy Corporation

STREET: 7 Skyline Drive

CITY: Hawthorne

STATE: NY

COUNTRY: USA

ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/457,646A

FILING DATE: 01-JUN-1995

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/457,205

FILING DATE: 01-JUN-1995

APPLICATION NUMBER: 08/258,261

FILING DATE: 08-Jun-1994

ATTORNEY/AGENT INFORMATION:

NAME: Elmer, James Scott

REGISTRATION NUMBER: 36,129

REFERENCE/DOCKET NUMBER: CGC 1506/CIP3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8614

TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 28958 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

US-08-457-646A-6

alignment\_scores:

Quality: 128.00 Length: 656

Ratio: 0.489 Gaps: 30

Percent Similarity: 39.939 Percent Identity: 20.427

alignment\_block:

US-09-316-048-6 x US-08-457-646A-6

Align seg 1/1 to: US-08-457-646A-6 from: 1 to: 28958

```
18 SerGlnIleLeuAsnLysAsnGlnSerLeuLeuSerGlnProLeu..... 32
   :::::::::::::::::::: :::::::::::::: ::::::::::
10574 ACATCGCTCCTCGGACGTGGACGCTCCCTCTTATTGACCGTCACCGGC 10623
   :::::::::::::::::::: :::::::::::::: ::::::::::
33 .....MetSerIleProSerThrThrSerSerLeuProSerGlnAsn 47
   :::::::::::::::::::: :::::::::::::: ::::::::::
10624 ACCAGGCTCGAGGCTCCGAGCTCGAGCGCGGCTGATCGAAGAACT 10673
   :::::::::::::::::::: :::::::::::::: ::::::::::
47 LagIATGProIleGlnAsnSerAlaLeuProSerAlaSerIleThrSer 63
   :::::::::::::::::::: :::::::::::::: ::::::::::
10674 CCGGAAACCGCTCCTGTTCTCGAGCGACCGAGCGGCTCTCGACATG 10723
   :::::::::::::::::::: :::::::::::::: ::::::::::
64 ThrSerAlaAlaLagIuSerIleThrProThrAlaGluLeuAsnAlaLe 80
   :::::::::::::::::::: :::::::::::::: ::::::::::
10724 GGATCGCTTCTCGCTCGAGGTCAGCCCC..... 10753
   :::::::::::::::::::: :::::::::::::: ::::::::::
80 uCysMetLysLeuGlyLysLysProMetTyrLysProValAsp..... 94
   :::::::::::::::::::: :::::::::::::: ::::::::::
10754 ....ATCCCGTGTACAGCTCGCCCTCCGCGAGACCTGCGAGCTCACC 10799
   :::::::::::::::::::: :::::::::::::: ::::::::::
95 .....ProTyrSer ArgMetGlnSerThrTyrAsnTyrAsnMetAr 108
   :::::::::::::::::::: :::::::::::::: ::::::::::
10800 GCTGATCCCGCTCGCTCGGCTCCATTGACGAGAGA..... 10838
   :::::::::::::::::::: :::::::::::::: ::::::::::
108 gGlyGlyAlaTyrProProArgTyrPheTyrProPheProValProProL 125
   :::::::::::::::::::: :::::::::::::: ::::::::::
10839 .....AGGCACCTCGC.....CCGCTGCTCTCTC 10863
   :::::::::::::::::::: :::::::::::::: ::::::::::
125 euleuTyrGlnValGluLeuSerValGlyGlyGlnGlnPheAsnGlyLys 141
   :::::::::::::::::::: :::::::::::::: ::::::::::
10864 TCCTG.....GGCGAGCTCTCTAC..... 10883
   :::::::::::::::::::: :::::::::::::: ::::::::::
142 GlySerThrArgGlnAlaAlaLysHisAspAlaAlaLysAlaLeuAr 158
   :::::::::::::::::::: :::::::::::::: ::::::::::
10884 ....CCGAGCCTCGCGCTCGAGTGAAGACTTCTTCGCGCTACAG 10927
   :::::::::::::::::::: :::::::::::::: ::::::::::
158 gIleLeuGln.....AsnGluProLeuProGluuArgLeu.... 169
   :::::::::::::::::::: :::::::::::::: ::::::::::
10928 CTCCCGCAAGGTCTCCCTCCCACTACCTCTCCACGAGACGGTTC 10977
   :::::::::::::::::::: :::::::::::::: ::::::::::
170 .....GluValAsnGlyArg..... 174
   :::::::::::::::::::: :::::::::::::: ::::::::::
10978 TGCTCGACGTCTCCACGAGACGAGTTCGAGCTGCGCTCCGACGGC 11027
   :::::::::::::::::::: :::::::::::::: ::::::::::
175 .....GluSerGluGluGlnAsnLeuAsnLys 183
   :::::::::::::::::::: :::::::::::::: ::::::::::
11028 TGACCTGCGCGACCAATCCCGTGTCTCGGCGCGCGCTCTCGCGCG 11077
   :::::::::::::::::::: :::::::::::::: ::::::::::
183 sSerGluIleSerGlnValPheGluIleAlaLeuLysArgAsnLeuProV 200
   :::::::::::::::::::: :::::::::::::: ::::::::::
11078 ACCGGGGTGGCTTCTCTTACAGGGCGGCTCTCCCTCGCAGACACCG 11127
   :::::::::::::::::::: :::::::::::::: ::::::::::
200 aL..... 200
   :::::::::::::::::::: :::::::::::::: ::::::::::
11128 TGCTCGAAGGCATGCCGTCTCGGCACACCACTCATCGGGCACCAGG 11177
   :::::::::::::::::::: :::::::::::::: ::::::::::
201 .....AsnPh 202
   :::::::::::::::::::: :::::::::::::: ::::::::::
11178 CTTTCTGAGCTCGCTCGACGTGCGCCACACCGCTCGGCTCGACACCG 11227
   :::::::::::::::::::: :::::::::::::: ::::::::::
202 eGluValAlaArgGluSerGly Pro.....ProHisMetLysAsn 215
   :::::::::::::::::::: :::::::::::::: ::::::::::
11228 TCGAAGAGCTACGCTCGAGGCGCTCTCGCTCTCCATCGCAGACACC 11277
   :::::::::::::::::::: :::::::::::::: ::::::::::
216 PheValThrLysValSerVal.GlyGluPheValGlyGluGlyLyl 232
   :::::::::::::::::::: :::::::::::::: ::::::::::
11278 GTCTCTCTCGATCTCGCTCGGCGCGTGGACGACGAGAGCAAGGGC 11327
   :::::::::::::::::::: :::::::::::::: ::::::::::
232 ysSerLysLysIleSerLysLysAsn..... 240
   :::::::::::::::::::: :::::::::::::: ::::::::::
11328 GCTCTCTTTCATAGCCGACAAAGAGACGCGCTTACAGATGCCCCCTGA 11377
```





```

476 leSerSerGlyHisValProHis.....GlyProLeuThrArg 488
      :::: :::: ||||| ||| |||
11914 TGGAGGGAGACTCCAAAGCCCCCACTTCACCGCCTGCCGCGCCGAGCGGCGCT 11963
      ||||| :::: ||||| :::: |||
489 ProSerGluGlnLeuAspTyrLeuSerArg.ValGlnGlyPheGlnValG 505
      ||||| :::: ||||| :::: |||
11964 CCTTCGCGGACACAGGCGGCCACGATCTCGCGCTCGACGCCGCCGCTCGCCC 12013
      ::||| ||||| :::: |||||
505 LuTyrLysAspPheProLysAsnAsnLysAsnGlnPheValSerLeuIle 521
      ::||| ||||| :::: |||||
12014 GCTACGCCGACCTCGCTGCCCTCCGAGC.....GCCCTCGAC 12051
      :::: :::: ||||| |||||
522 AsnCysSerSerGlnPro.....ProLeuIleSerHisG1 533
      :::: :::: ||||| |||||
12052 CAGGGCGCTTCGCGCTCCGCGCTCGTGTGCGGCCCTTCATCGATCGAGCC 12101
      ||||| :::: ||||| :::: |||
533 ylleGlyLysAspValGluSerCysHisAspMetAlaAlaLeuAsnIleL 550
      ||||| ||||| :::: ||||| :::: |||
12102 GGCAGGCGACCTCGTCCGAGCGCCCGAGGCGCACCGCGCTCGCACTCG 12151
      |||||
550 euLysLeu 552
      |||
12152 CCCTCTTG 12159

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2001, 00:30:32 ; Search time 11371 Seconds

(without alignments)  
1760.655 Million cell updates/sec

Title: us-09-316-048-9

Perfect score: 2857

Sequence: 1 cggcgcgcgccctccccg.....ccaataatgtaactaa 2857

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7991742 segs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: em\_estda:\*  
42: em\_estfun:\*  
43: em\_esthum1:\*

44: em\_esthum2:\*  
45: em\_esthum3:\*  
46: em\_esthum4:\*  
47: em\_esthum5:\*  
48: em\_esthum6:\*  
49: em\_esthum7:\*  
50: em\_esthum8:\*  
51: em\_esthum9:\*  
52: em\_esthum10:\*  
53: em\_esthum11:\*  
54: em\_esthum12:\*  
55: em\_esthum13:\*  
56: em\_esthum14:\*  
57: em\_esthum15:\*  
58: em\_esthum16:\*  
59: em\_esthum17:\*  
60: em\_esthum18:\*  
61: em\_esthum19:\*  
62: em\_esthum20:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estcov1:\*  
68: em\_estcov2:\*  
69: em\_estp11:\*  
70: em\_estp12:\*  
71: em\_estp13:\*  
72: em\_estp14:\*  
73: em\_estp15:\*  
74: em\_estro1:\*  
75: em\_estro2:\*  
76: em\_estro3:\*  
77: em\_estro4:\*  
78: em\_estro5:\*  
79: em\_estro6:\*  
80: em\_estro7:\*  
81: em\_estro8:\*  
82: em\_estro9:\*  
83: em\_estro10:\*  
84: em\_estro11:\*  
85: em\_estro12:\*  
86: em\_estro13:\*  
87: gb\_est41:\*  
88: gb\_est42:\*  
89: gb\_est43:\*  
90: gb\_est44:\*  
91: gb\_est45:\*  
92: gb\_est46:\*  
93: gb\_est47:\*  
94: gb\_est48:\*  
95: gb\_est49:\*  
96: gb\_est50:\*  
97: gb\_est51:\*  
98: gb\_est52:\*  
99: gb\_est53:\*  
100: gb\_est54:\*  
101: gb\_est55:\*  
102: gb\_est56:\*  
103: gb\_est57:\*  
104: gb\_est58:\*  
105: gb\_est59:\*  
106: gb\_est60:\*  
107: gb\_est70:\*  
108: gb\_est71:\*  
109: gb\_est72:\*  
110: gb\_est73:\*  
111: gb\_est74:\*  
112: em\_esthum21:\*  
113: em\_esthum22:\*  
114: em\_esthum23:\*  
115: em\_estom1:\*  
116: em\_estom2:\*

```

117: em_estp16:*
118: em_estp17:*
119: em_estp18:*
120: em_estp19:*
121: em_estp20:*
122: em_estp21:*
123: em_estp22:*
124: em_estp23:*
125: em_estp24:*
126: em_estp25:*
127: em_estp26:*
128: em_estp27:*
129: em_estp28:*
130: em_estp29:*
131: em_estp30:*
132: em_estp31:*
133: em_estp32:*
134: em_estp33:*
135: em_estp34:*
136: em_estp35:*
137: em_estp36:*
138: em_estp37:*
139: em_estp38:*
140: em_estp39:*
141: em_estp40:*
142: em_estp41:*
143: em_estp42:*
144: em_estp43:*
145: em_estp44:*
146: em_estp45:*
147: em_estp46:*
148: em_estp47:*
149: em_estp48:*
150: em_estp49:*
151: em_estp50:*
152: em_estp51:*
153: em_estp52:*
154: em_estp53:*
155: em_estp54:*
156: em_estp55:*
157: em_estp56:*
158: em_estp57:*
159: em_estp58:*
160: em_estp59:*
161: em_estp60:*
162: em_estp61:*
163: em_estp62:*
164: em_estp63:*
165: em_estp64:*
166: em_estp65:*
167: em_estp66:*
168: em_estp67:*
169: em_estp68:*
170: em_estp69:*
171: em_estp70:*
172: em_estp71:*
173: em_estp72:*
174: em_estp73:*
175: em_estp74:*
176: em_estp75:*
177: em_estp76:*
178: em_estp77:*
179: em_estp78:*
180: em_estp79:*
181: em_estp80:*
182: em_estp81:*
183: em_estp82:*
184: em_estp83:*
185: em_estp84:*
186: em_estp85:*
187: em_estp86:*
188: em_estp87:*
189: em_estp88:*

```

```

190: gb_gss25:*
191: gb_gss26:*
192: gb_gss27:*
193: gb_gss28:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	665	23.3	688	92	AM549911 L0057C11-
2	596.6	20.9	633	89	AM320695 uc022b10.y
3	582.6	20.4	1018	136	BE867390 601442470
4	554.6	19.4	725	106	BE306047 601101723
5	554.4	19.4	624	110	BE623404 uc91f11.x
6	528.8	18.5	677	92	AM555746 L0259A06-
7	528.6	18.4	686	91	AM537709 G0120F09-
8	525.8	18.4	564	2	AA120063 mp88b08.r
9	517.2	18.1	522	139	BF016149 yf01e08.y
10	511	17.9	522	2	AA122533 mp5c04.r
11	508.4	17.8	752	137	BE894295 601437666
12	507.8	17.8	511	92	AM542549 C0158H10-
13	502.2	17.6	616	111	BE674157 7d76d04.x
14	499.4	17.5	520	147	WB9457 m73e03.r1
15	486	17.0	497	92	AM541590 C0146H10-
16	467	16.3	467	106	BE286356 601097825
17	463.4	16.2	609	136	BE810330 MR4-PT005
18	462.4	16.2	504	40	AM141293 EST291334
19	441.2	15.4	481	4	AA240112 my22a02.r
20	440	15.4	452	105	BE198783 ug82e04.y
21	435.2	15.2	594	40	AM162002 au72c03.x
22	435	15.2	528	28	AU021763 AU021763
23	434	15.2	686	110	BE613972 601504093
24	426	14.9	464	18	A1286843 uc55d11.r
25	426	14.9	1801	146	U69197 U69197 S0aR
26	422	14.8	724	97	AM952360 EST364430
27	415.8	14.6	442	13	AA880492 vx40h06.r
28	414.8	14.5	852	137	BE872137 601446238
29	414.6	14.5	829	106	BE300155 600944602
30	414.2	14.5	564	40	AM163206 au97f05.x
31	406.2	14.2	446	136	BE865252 ut-m-BH0-
32	401	14.0	401	2	AA106767 m186f12.r
33	398.8	14.0	615	134	BE082712 RC2-BT064
34	395.6	13.8	424	16	A1155343 uc88c03.r
35	395	13.8	425	16	A1152490 uc83e09.r
36	387.2	13.6	673	135	BE781223 601469159
37	386.8	13.5	573	3	AA206573 zq51f08.r
38	386	13.5	398	147	W99833 m926f09.r1
39	384.4	13.5	524	96	AM877286 MR4-PT005
40	379	13.3	384	2	AA124161 mp94g04.r
41	377	13.2	377	2	AA073987 mm95d04.r
42	376.6	13.2	391	29	AV027862 AV027862
43	376.2	13.2	866	136	BE868328 601443867
44	373.8	13.1	475	4	AA270608 va6b02.r
45	372.6	13.0	543	134	BE089861 RC5-BT070

## ALIGNMENTS

```

RESULT 1
AM549911/c 688 bp mRNA EST 31-AUG-2000
LOCUS L0057C11-3 Mouse fl2.5 Female Mesonephros and Gonads cDNA Library
DEFINITION Mus musculus cDNA clone L0057C11 3', mRNA sequence.
ACCESSION AM549911
VERSION AM549911.1 GI:7195339
KEYWORDS EST, house mouse.
SOURCE

```



ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 688)  
AUTHORS Tanaka,T.S., Jardard,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac,M.J., Pantano,S., Sano,Y., Piao,X., Nagaraja,R., Dol,H., Wood,W.H., III, Becker,K.G. and Ko,M.S.H.  
TITLE Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)  
MEDLINE 20381348  
COMMENT Contact: George J. Kargul  
Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: cdba@sun.grc.nia.nih.gov  
Plate: L0057 row: C column: 11  
Seq primer: -21M3 Forward  
High quality sequence stop: 688  
POLYA=Yes.

FEATURES  
source  
1. 688  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="L0057C11"  
/clone\_1ib="Mouse E12.5 Female Mesonephros and Gonads cDNA library"  
/sex="female"  
/dev\_stage="12.5dpc"  
/lab\_host="DH10B"  
/note="Vector: pSPORT1 (Gibco/BRL Life Technology);  
Site:1: Salt: Site:2: NotI: Total RNAs were extracted from 2 Mesonephros. The double-stranded cDNA was synthesized by Gibco's kit with an Oligo(dT) primer [NotI primer-adaptor: from GibcoBRL]  
15'-pGACTAGTTCATGATCGGAGCGCGCCCTTTT-3' from 3.42ug of total RNA. The double-stranded cDNAs were treated with T4 DNA polymerase and purified by ethanol-precipitation. The cDNAs were ligated to lona-linker L1-SalI (include SalI sequence). The cDNAs were purified by phenol/chloroform and separated from free linkers by Centricon 100. Then, cDNAs were amplified by long-range high fidelity PCR using Takara's Ex Taq polymerase. Then, the cDNAs were purified by phenol/chloroform and by Centricon 100. The cDNAs were digested with SalI and NotI enzymes. Then, the cDNAs were size selected by Gibco's Size Fractionation Column. The cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by chemical method. The library was constructed by Xiaohong Wang."

BASE COUNT 209 a 144 c 162 g 173 t  
ORIGIN

Query Match 23.3%; Score 665; DB 92; Length 688;  
Best Local Similarity 99.7%; Pred. No. 1,1e-173;  
Matches 667; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Db 2168 cacatgactgactgactgcgcacacagtagcaccactgtaaccactgtcttcttc-a 2226  
|||||  
Db 688 CACATGACGACGACAG-TCGGCCACACGTAACGACACTGTAAACCGCTTCTTCMA 630  
|||||  
Qy 2227 gttcatttttttcccttgatgtatatacaataatttcatttcagttcctgtg 2286  
|||||  
Db 629 GTTTCATTTTTCCTTGTGATTGATACACACTATATTTTCATTTCCCTTGTGCG 570  
|||||  
Qy 2287 tgtctacttaccagagtttagaactgcatgcatgtaacagcgaagatcacagccc 2346  
|||||  
Db 569 TGCTACTTACCTACGACAGTTTAGAAACTGTCACTCATGTACTGCGCAAGGATACAGCCC 510  
|||||  
Qy 2347 ggttgaggtagcattctgtcctctgtgctgtgaacagttctgtgaattacacacagaat 2406

|||||  
Db 509 GGTGGGTGGCATTTCTGTGCTGCTGGCTGGCGAAGAGTTGGGAATTAACACAGAAAT 450  
|||||  
Qy 2407 ccttgactccctccctcttgataaattgagacagcttaaggactttaaacttagatcaa 2466  
|||||  
Db 449 CTTGTACTCCCTCCCTCTGTATTAATTTGACAGCTTTAGACTTTTAACCTTAGATCA 330  
|||||  
Qy 2467 aagatagtctctttttaacttattttaaggagacagactttaaataatgagccctgact 2526  
|||||  
Db 389 AAGATATGTGCTTTTAATTATTTTAAGAGAGAGACTTTAAATAGAGCCCTGACCT 330  
|||||  
Qy 2527 taaccattataacagaatttgcataaaggagtgttcttgaggaggtagcttttttta 2586  
|||||  
Db 329 TTACCCATTATATACAAATTTGTCAAAAGAGAGCTTTCTTGAGAGAGTACTTTTTTTTA 270  
|||||  
Qy 2587 ccacactacagagacattacctgtgagcccaagaagactacagagctgtgtccccaagagggc 2646  
|||||  
Db 269 CCACACTACAGGACATTAAGCTGTAGGCCCGAGAGACTACAGGCTGGGTCTCTAGAGGGC 210  
|||||  
Qy 2647 ccaatacagtcgaattccaaccttaagtcgcggggaagtgacaggttccctgtgtgct 2706  
|||||  
Db 209 CCAATACAGTCATTTCCAACTCTAAGTGGGGAAGGTACAGGTTTCTGTGCTGGGT 150  
|||||  
Qy 2707 gtgcacaggggcagcagagtgctgagcctggtgggaagagcattgtgctctagtgacgc 2766  
|||||  
Db 149 GTGCACAGGGGCGAGCAGGTCACTGGCCCTGGGGAAGAGCATTTGTGGCTCTAGTCAGC 90  
|||||  
Qy 2767 cctggttccactctgtgttagctgtgaaccttcccaactatggaataaagtaactcac 2826  
|||||  
Db 89 CCGTCTTCACCTCTTGTTAGCTGGGAACCTTCCCACTATGGAATATAGTAACCTCAC 30  
|||||  
Qy 2827 ttctctgtaccataaatgtgaataact 2855  
|||||  
Db 29 TTTCCTTTGTACCAATTAATGTAATACT 1

RESULT 2  
AM320695 633 bp mRNA EST 25-JAN-2000  
LOCUS uc022h10.y1 NCL\_CGAP\_Mam6 Mus musculus cDNA clone IMAGE:2609851 5'  
DEFINITION Similar to TR:095793 095793 STAU6N PROTEIN.; mRNA sequence.  
ACCESSION AM320695  
VERSION AM320695.1 GI:6750239  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 633)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue procurement: Jeffrey Green M.D.  
cDNA library Preparation: Life Technologies, Inc.  
cDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/dbref/image/image.html](http://www.bio.llnl.gov/dbref/image/image.html)

FEATURES  
source  
1. 633  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:2609851"

```

/clone_1lb="NCL.CGAP.Man6"
/sex="Female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: PCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"

```

BASE COUNT 176 a 167 c 163 g 127 t

Query Match 20.9%; Score 596.6; DB 89; Length 633;  
 Best Local Similarity 97.7%; Pred. No. 1.2e-154;  
 Matches 605; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

```

Oy 1123 ccagagcgacgctgccaagcagcactcaatcagaagaagaagcctcagtaagaagc 1182
    |||||||
Db 14 CCCAGGCGCAGCCTGCCAAGCCAGCATCAATCAGAAAGAACTCCAGTAAGAAAC 73
    |||||||
Oy 1183 cagaggaaggaagaagtaagtttttgaacctagccctgggagtgaaatggaacta 1242
    |||||||
Db 74 CAGGAGACGGAAGAAAGTAAGTTTGAACCTAGCCCTGGGATGAAATGGAACATA 133
    |||||||
Oy 1243 gtaacaagaagaagatcagatgacctatcttagccatcagcagctgcaagctgga 1302
    |||||||
Db 134 GTAAACAAGACAGAGAGATTGACAGATGCTTATCTTAGCATACAGAGCTGCCAGCTGGA 193
    |||||||
Oy 1303 ttctcccatggtgcggaagttgcccaggtgctgcggggttagtcaagaagacacacaca 1362
    |||||||
Db 194 TTCTCCCATGGTGCCTGCGGAAGTTGCCAGGCTGTGGGGTTAGTCAAGACACCAACCA 253
    |||||||
Oy 1363 aagatttaacagagcagctcccaaatctcgcgaagcgaagtaactgcatgataagccc 1422
    |||||||
Db 254 AAGATTTCACCGAGGCGAGCTCCAAATCCGCCAAGCAACGGTAACCTGCCATGATAGCCC 313
    |||||||
Oy 1423 gagaattgttgcgaggggacactgcgccaagcagcagacacatttaagaagtaacatct 1482
    |||||||
Db 314 GAGAGTTGTTGTAAGGGGGACCTCGCCACAGCCGAGACATTTTAAAGATTAACATCT 373
    |||||||
Oy 1483 ctccagagcagtcaccacatggaacctgcagctagaccctctgaagcaactgtaactctt 1542
    |||||||
Db 374 CTTCAGGCGACCTACCCGATGACCTCGCAGTAAGACCTCTGAGCAACCTGTAACCTTT 433
    |||||||
Oy 1543 ccaagagcccaagagattccaagttgaatacaagatttcccaagaacaaagaagcagct 1602
    |||||||
Db 434 CCAGAGCCCAAGGATTCAGAGTTGAATCAAAAGATGTTGCCAAGACGAAACGAGAGT 493
    |||||||
Oy 1603 gtgtatctctcaacatgctcctcacagccgctcctgcagtcagtcagtcagtcagcag 1662
    |||||||
Db 494 GTGTATCTCTCATCAATGCTCTCTACAGCCGCTCTGTCAGTATGCGATCGGCAAG 553
    |||||||
Oy 1663 atgtggaagtcctgcatatgataatgctgacatgaacatttaagctgctgctgagttgg 1722
    |||||||
Db 554 ATGTGAAGCTCTGATGATATGCTGCTGACATGACATTTGAAGCTGCTGTGTGAGTTGG 613
    |||||||
Oy 1723 accaacaagaagcagaagat 1741
    |||||||
Db 614 ACTCACAGAACACAGATAT 632
    |||||||

```

RESULT 3  
 BEB67390 1018 bp mRNA EST 27-SEP-2000  
 LOCUS 60144470F1 NIH\_MGC\_65 Homo sapiens cDNA clone IMAGE:3846689 5',  
 DEFINITION mRNA sequence.  
 BEB67390  
 ACCESSION BEB67390.1 GI:10316166  
 VERSION BEB67390.1  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
 1 (bases 1 to 1018)  
 AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
 Plate: LCM546 row: 1 column: 18  
 High quality sequence stop: 654.

FEATURES  
 source  
 Location/Qualifiers  
 1..1018  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3846689"  
 /clone\_1lb="NIH\_MGC\_65"  
 /tissue\_type="adenocarcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: colon; Vector: PCMV-SPORT6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.8 kb. Library constructed by Life  
 Technologies."

BASE COUNT 295 a 267 c 262 g 194 t  
 ORIGIN

Query Match 20.4%; Score 582.6; DB 136; Length 1018;  
 Best Local Similarity 81.8%; Pred. No. 1.1e-150;  
 Matches 735; Conservative 0; Mismatches 154; Indels 10; Gaps 5;

```

Oy 925 ttgcacagatccagcagcgaaggaagaaggaagcagagctacatgctcttaccagaac 984
    |||||||
Db 1 TGCCCATGATCCAGCAGGCAAAAAGAGAGAGACCAAGATACAGCTCTCCACAGAGC 60
    |||||||
Oy 985 gaggcttccacagctgcagaggaagttgtgtagcaggtgaaggttggtgacacactcag 1044
    |||||||
Db 61 GAGGCTCTCCCGCGCGGAGGAGTTGTGATGACAGTGAAGTTGGAACACCACTGCAG 120
    |||||||
Oy 1045 aaggagtggttaccataagaagtggtgccaagcgtgaatgctgtagaacatgctggaga 1104
    |||||||
Db 121 AAGGAGCGGCGACCAACAAGAGGTGGCCAAAGCCATGACAGCGAAGACATGCTGGAGA 180
    |||||||
Oy 1105 tcttgaggttcaaatctcccccagcagcagcctgcgaagcagcagcactcaatcagaaga 1164
    |||||||
Db 181 TCCTTGTTTCAAAATCCCGCAGCGGACCCACCAAAACCCGCACTCAAGTCAAGAGAGA 240
    |||||||
Oy 1165 agactccagtaagaagaacaggaagcgaagaagtaagcttlttgaacctagccctg 1224
    |||||||
Db 241 AGACACCCATTAAGAAACACAGGAGATGAGAAAGTAACCTTTTGAACCTGACTGCTG 300
    |||||||
Oy 1225 gggatgaataatggaactgtaacaagaagcaggttcaagaatgcttacttaccatc 1284
    |||||||
Db 301 GGGATGAATAATGGAGCTAGTAATAAGAGATGAGTTGAGGATGCTTATTAATCTATC 360
    |||||||
Oy 1285 agcagctgcagctggaatttcccccagtggtgcgggaagttgccaagcttgcgggtta 1344
    |||||||
Db 361 AGCAGCTCTGCTGGAATTTCTCCATGAGTGCCTCCGAGATGCCAGGCTGTAGAGATTA 420
    |||||||
Oy 1345 gtcaagagacacacacaaagattccaccaggaagcagctccaaatcctcgaagcgaag 1404
    |||||||
Db 421 GTCAAGACATCACACCAAAAGATTTTACAGGGCAGCTCCGATCTCTCCAAAGGCAAGG 480
    |||||||
Oy 1405 taactgcatgataagcccgagagttgtgtatcaggggacactcgccacagccgaagaca 1464
    |||||||
Db 481 TAACTGCATGATACCCGAGAGTTGTGTATGAGGGGACCTCGGCCACAGCCGAGACCA 540
    |||||||

```

Query Match	19.4%;	Score 554.6;	DB 106;	Length 725;
Best Local Similarity	92.7%;	Pred. NO. 5.8e-143;		

Email: [Robert\\_Strausberg@nih.gov](mailto:Robert_Strausberg@nih.gov)  
This clone is available royalty-free through LNL ; contact the

IMAGE Consortium (Info@image.llnl.gov) for further information.  
MGI:1079441

High quality sequence stop: 445.

## FEATURES

Location/Qualifiers

1..624

/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3369837"  
/clone\_lib="Soares.mammary-gland\_NMLMG"  
/sex="female (lactating)"  
/tissue\_type="mammary gland"  
/lab\_host="DHI0B"  
/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker: 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 153 a 138 c 155 g 178 t

## ORIGIN

Query Match 19.4%; Score 554.4; DB 110; Length 624;  
Best Local Similarity 96.2%; Pred. No. 6.2e-143;  
Matches 600; Conservative 0; Mismatches 21; Indels 3; Gaps 3;

QY 1653 atcggcaagaagtgtgagtcctgtcatgatgtgtgcatacaatttaagctgtc 1712  
DB 624 ATCGGCACGAGATGTGAAGCTTGTATGATATGCGACGACTGAACTTTTACAGCTGTG 565  
QY 1713 tctgagttgaccaacagagacagatgcagaagagagaatgaccagtttcacg 1772  
DB 564 TATGAGTGGACCTACAGAGCTCAGAGATGCCAGAAACAGAGATGGCCAGTTTCAGCG 505  
QY 1773 tgcggagagtgctgaacctctctgcgcacaaacacatlaaaaccacacatatactg 1832  
DB 504 TCGCGAGGTGCTGAAACCTTTCTGCGCCCAACCTTTTAAAGCCACATATATACG 445  
QY 1833 aaaaactcgaagacgtcttgaaaatttgaaatctgaactccagtgaggccaagaca 1892  
DB 444 AAAAATACGAGAACTGCTTTGAAATTTGAAATATCTGATACCTCCAGTGGGCAAGACA 385  
QY 1893 tgggtgataaaaaatctggcaaaagacagaagaacttcagtgtagccctgtgtgtc 1952  
DB 384 TGGTGGATTAATAATGTGGCAAAAGACACAGAAACTTCAGGTGTAGCCCTGTGTGTCT 325  
QY 1953 -ggcggtagtgaatgctgtgtctgtccatccacagagaagaaccagcccaagc 2011  
DB 324 GGGCGGCTGTGATGATGCTGTGCTGTGCATTCATCCAGACAGAAACCCGCAAGCC 265  
QY 2012 ctccagttctgttttgcacgtgacaaagagacagacccaatttcacatgtctgtctc 2071  
DB 264 CTCACGTTCTGTTTTTGGCATCTGACAAAGAGACACACCAATTCATGCTGGCTTCT 205  
QY 2072 tcaagtaacttgaaaaaccggagacagccaccacagagagccttataggcgcccgagac 2131  
DB 204 TTAGATACCTTTTAATAAACCCGACACACACACAGAGAGGCTTTATAGGCGCCGGAGC 145  
QY 2132 taaacggaccagagaagaagcagtgctctcactactgacatgacagcagctccgaca 2191  
DB 144 TAAAGCACACAGAGAAA-ACCAGTGTCTCTACGTGCACATGACATGACCTCCGCCA 86  
QY 2192 caagtagcaacacgtgaacactgtctctctc-agttcatcttttctccttgatga 2250  
DB 85 CACGTAGCACCCACTGTAAACACACTGCTTCTCTCAAGTTTCAATTTTTCCTTTATTA 26  
QY 2251 tacaacacataatcttcaattca 2274  
DB 25 TACAACACTATAATTTTCAATTTCA 2

RESULT 6  
AM555746/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES  
source

AM555746 677 bp. mRNA EST 31-AUG-2000  
LOCUS L0259A06-3 Mouse Newborn Ovary cDNA Library Mus musculus cDNA clone  
DEFINITION L0259A06-3', mRNA sequence.  
ACCESSION AM555746  
VERSION AM555746.1 GI:7201169  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 677)  
Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grabovac  
'M.J., Pantano,S., Sano,Y., Piao,Y., Nagasaja,R., Dol,H., Wood,W.H.  
III, Becker,K.G. and Ko,M.S.H.  
Genome-wide expression profiling of mid-gestation placenta and  
embryo using a 15,000 mouse developmental cDNA microarray  
Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)  
20381348  
Contact: George J. Kargul  
Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: cdna@igsun.gic.nia.nih.gov  
Plate: L0259 row: A column: 06  
Seq primer: -21M13 Forward  
High quality sequence stop: 677.  
POLY-A-Tes.

Location/Qualifiers

1..677

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="L0259A06"

/clone\_lib="Mouse Newborn Ovary cDNA Library"

/sex="female"

/dev\_stage="Newborn Ovary"

/lab\_host="DHI0B"

/note="Vector: pSPOR1 (Gibco/BRL Life Technology);

Site\_1: SalI; Site\_2: NotI; Total RNAs were extracted from

7 Newborn Ovary. The double-stranded cDNA was synthesized

by Gibco's kit with an Oligo(dT) primer [NotI

primer-adaptor from GibcoBRL]

[5'-pGCTACTTCTAGATCGGAGCGGCCCTTTT-3'] from

2.50ug of total RNA. The double-stranded cDNAs were

treated with T4 DNA polymerase and purified by

ethanol-precipitation. The cDNAs were ligated to

lone-linker L1-Sal3 (include SalI sequence). The cDNAs

were purified by phenol/chloroform and separated from

free linkers by Centricon 100. Then, cDNAs were amplified

by long-range high fidelity PCR using Takara's Ex Taq

polymerase. Then, the cDNAs were purified by

phenol/chloroform and by Centricon 100. The cDNAs were

digested with SalI and NotI enzymes. Then, the cDNAs were

size selected by Gibco's Size Fractionation Column. The

cDNAs were cloned into SalI/NotI site of pSPOR1 plasmid

vector. The DHI0B E. coli host was transformed with the

ligation mixture by chemical method. The library was

constructed by Xiaohong Wang and Yulan Piao."

BASE COUNT 162 a 158 c 163 g 194 t

## ORIGIN

Query Match 18.5%; Score 529.8; DB 92; Length 677;  
Best Local Similarity 92.6%; Pred. No. 4.5e-136;  
Matches 629; Conservative 0; Mismatches 2; Indels 48; Gaps 5;

QY 1641 gtcaatcatgcatgcgaagaagtgtgagtcctgtcatgatgtgtgcacgtgaacatt 1700  
DB 677 GTCAATCATGCGATCGGACAGGATGTGAGTCTCTCATGATGATGCTGCACATGAACAT 618  
QY 1701 ttaagctgctgtctgagttgaccaacagagcacagatgccaagaacgaaatga 1760

D	617	TTAAAGCGTGTGATATTGGACACAGAGACACAGATGCCAAGAACGGAATATGA	558
OY	1761	ccagttcagc-----gtgcg	1776
D	557	CCAGTTTCAGCGTGATGCACACATGCAGAAAGCAATGGCAGTGTGGGTGAGCACAGTGGG	498
OY	1777	ggaggctgtgaaccttcttcgtgccacaacattataaacccaatataacttgaaaa	1836
D	497	GGAGGTGCTGAACCCTTTCTGTGGCCCAAACCATWTTAAAAACCAACATATAATACGAAA	438
OY	1837	tactagaacacgtccttgaanaatttgaaatcatctgaatactccagltgggccaagatatgt	1896
D	437	TACTAGAACACTGCTTTGAAAATTBTGGAATATCTGATTAACCTCAGTGGCCAGACATGCT	378
OY	1897	ggataaaaatlylgcaaaagcacagaanaacttcagltgtagccccgttgtlgt-99c	1955
D	377	GGATMAAATGTCGCAAAAGACAGACAAGAAACTTCAGTGGTGTGGCTGTGCTGGGC	318
OY	1956	ggtctagtatgatctgtgtctgtccatccatccagacaagaaccagcccagcctcc	2015
D	317	GGCATGATATGATGCTGTGCTCTGTCCATCCACAGAAAGAAACCAAGCCCCAACCCCTCC	258
OY	2016	agctctgtcttctgacgtctgacaagagagacacacatctcatctgtcgcttcttaag	2075
D	257	AGTTCTGTTTTTTCATCGTGTGACAAAGAGAGCACACCAATTCATGCTGCTCTTCAG	198
OY	2076	atacttgaaaaacccggacagccacacacagagagcceltatagagcccgagactaa	2135
D	197	ATACTTTGAAAAACCCGGACAGCCACACACAGAGAGCGCTTATAGGGGCCCGGAGACTAA	138
OY	2136	cggaccagagaagaagcagtgctctctactctgacatgactgaactgaactccgcacagc	2195
D	137	CGCACCAAGAGAAA-ACCAAGTCTTCTCAATGACATGACTGACTCAG-TCCGCCACACG	80
OY	2196	tagccacctbaacctgctctcttc-agtttcgttttctccttgatgataca	2254
D	79	TAGCACCACTGTAACACTGCTTCTCTCAAGTTTCATTTTTTCTTGATATATACA	20
OY	2255	aacactaatcttcattcc 2273	
D	19	ACACTATAATTTTCAATTTC 1	
RESULT	7		
LOCUS	AW537709/c		
DEFINITION	G0120F09-3 Mouse E7.5 Embryonic Portion cdNA Library Mus musculus		
ACCESSION	CDNA clone G0120F09 3', mRNA sequence.		
VERSION	AW537709		
KEYWORDS	AW537709.1 GI:7180222		
SOURCE	EST.		
ORGANISM	house mouse. Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 686)		
AUTHORS	Tanaka,T.S., Jaradate,S.A., Lim,M.K., Kargui,G.J., Wang,X., Grabovac, M.J., Pantano,S., Sano,Y., Piao,Y., Negaraja,R., Dol,H., Wood,W.H. III, Becker,K.G. and Ko,M.S.H.		
TITLE	Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)		
MEDLINE	20381348		
COMMENT	Contact: George J. Kargui Laboratory of Genetics National Institute on Aging/National Institutes of Health 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA Email: cdna@gsun-grc.nia.nih.gov Plate: G0120 row: F column: 09 Seq primer: -2IM13 Forward High quality sequence stop: 686 POLYA=yes.		

FEATURES  
SOURCE

Location/Qualifiers

```
/organism="Mus musculus"
```

```

/db_xref="taxon:10090"

```

```
/clone_lib="Mouse E7.5 E
```

```
/dev_stage="7.5dpc Embryo"
```

```
/note="Vector: pSPORT1 (Gi
```

6 Embryo. The double-stranded cDNA was synthesized by

gibco s kit with an oligo(dT) primer (Noli primer-adaptor  
from gibcoBRL]

[3'-pGACTAGTCTAGATCGGAGCGCCCTTTTCTTTT-3']  
from 0.51  $\mu$ g of mRNA. The double-stranded cDNAs were

ethanol-precipitation. The cDNAs were ligated to

were purified by phenol/chloroform and separated from

free linkers by Centricon 100. Then, cDNAs were amplified by long-range high fidelity PCR using Takara's Ex Taq

polymerase. Then, the cDNAs were purified by phenol/chloroform and by Centricon 100. The cDNAs were

digested with SalI and NotI enzymes. Then, the cDNAs were

CDNAs were cloned into SalI/NotI site of pSPORT1 plasmid

ligation mixture by chemical method. The library was

a	139	c	150	g	183	t
---	-----	---	-----	---	-----	---

18.48; Score 526.6; DB 91; Length 686;

```

    33.18; Freq. NO, 3.06 100;
servative 0; Mismatches 4; Indels 1; Gaps 1;

```

tgtaactgccaagatcacacagcccggttgqgtgqcatctgtgcctctg 2372

TGTAACGGCAAGGATCACAGCCCGGTGGGTGGCAATTCTGTGCTCTGGG 627

caattctgaataaccacgaatccttaactcctgcccttataaa 2432

CAGTTCTGGAATTACCAACAGATCCTTGACCTCCCTGCCCCCTGTATAAA 567

2492. ttttaaacctttatgatacaaaacatattttcttttaaccttatt

507

[illegible]

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841

100

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

[illegible][illegible]

70

GGTACACAGTTTCCTGGTGCTGGGTGTCACACAGGGGCAGGCAGGTCAGCTG 208

gagcattgtgctccctagtcagccctgtctccactcttggctttagctgg 2/92

GAGCATTGTGGCTCCTAGTGCAGCCCTGCTTCCACTCTTGGTATTAGCTGG 208

ctcatggaatataagtaaactcacttcttctgtcaccataaatgtgaat 2852

Db	207	AAACCTCCACGATCATGATATTAAGTAACACACACCTCTTCTTGTCACCAATAAATGTAT	148
Qy	2853	actaa 2857	
Db	147	ACTA 143	
RESULT	8		
LOCUS	AA120063	564 bp mRNA	EST 17-FEB-1997
DEFINITION	mp88b08.1 Soares_thymus_2nbmt Mus musculus cDNA clone IMAGE:576279		
ACCESSION	AA120063		
VERSION	AA120063.1	GI:1677763	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.		
TITLE	The WashU-HHMI Mouse EST Project		
JOURNAL	Unpublished (1996)		
COMMENT	Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:350927 Seq primer: -28M13 rev2 from Amersham High quality sequence stop: 488. Location/Qualifiers 1..564 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone_image="576279" /clone_lib="Soares_thymus_2nbmt" /sex="male" /tissue_type="Thymus" /dev_stage="4 weeks" /lab_host="DH10B" /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCACTGAGGAGGAGGAGCGGCGGTTTTTTTTTTTTTTTTTTTTT 3'] double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Bertrand Jordan. Library went through two rounds of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."		
BASE COUNT	150 a 142 c 124 g 148 t		
ORIGIN			
Query Match	18.4%; Score 525.8; DB 2; Length 564;		
Best Local Similarity	99.1%; Pred. No. 5.4e-135;		
Matches 560; Conservative 0; Mismatches 2; Indels 3; Gaps 3;			
Qy	1831	tgaatactgagaactgcttgaataatctggaatatctgataactcagtgaggcaaga 1890	
Db	1	TGAAATACTGAGACGCTTTGAAATAATTTGGAAATATCTGATTAACCTCAGTGCGCCAGAGA 60	
Qy	1891	catggtgataaaaaatctgccaagaagcaagaactcagtggtgtagccctggtgtg 1950	

Db	61	CATGGTGAATAAATGATGGCAAGACACAAGAANAATTTCAGGTGGTAGCCCTGGTTTGTC	120
OY	1951	ct-aggcgctagtgatgatgtcgtgcttgcacatccacgaagaacaagcccac	2009
Db	121	CTGGCGCGCTAGTGTATGATGCTGTGCCTGTGCATCCATCCAGACAGAAACACAGCCCCAAC	180
OY	2010	ggcttcagttctgttttttgacatcgtgcaaaagagagacacagccaattctcatgtctgacct	2069
Db	181	GCCCTCAGCTTCTGTTTTTGTGCAATCGGAACAAAAGAGACACAGCAATTTCTCATGCTGGCTT	240
OY	2070	cttcagataacttggaaaaaccggacgaacacacacagaagagccttatagcgccccga	2129
Db	241	CTTGAGATACTTTGAAAAACC GGACACCCACACACACAGAGAGGCCCTTAATAGGGCCCCCGGA	300
OY	2130	gctaaaeggaccagaagaagagccagtgctcttccctaactygcacatgactgtaactcagctccgc	2189
Db	301	GCTAAACCCACACAGAGAAAA -CCAGTGCCTTCCCTCACTACATGACTGACTCAGCTCCGC	359
OY	2190	cacacgtgtagcaaacactgytaaacacgcttctctctc -agttaatttttttcccttgaat	2248
Db	360	CACACGTAGCACCACTGTAACCACTGCTTCTCTTCAAGTTTCATTTTTCCTTGTGATT	419
OY	2249	gatacaacacataatttcaatttcagttcccttagtcgtgcttactactaacagagttta	2308
Db	420	GATACACACTATATTTTTCATTTCAGTTCTCTTAAGTCTGCTACTTACCTACAGAGCTTTA	479
OY	2309	gaaactgtcagtcatalgtactgycgaagatacacagcccagttggtygcatctgtgacct	2368
Db	480	GAACCTGTACATGATGTAAGTCGCAAGAGATACACAGCCGGTGGGTGCATTCGTGCT	539
OY	2369	ctgagcttgagctgtaaacagttctggaa	2393
Db	540	CTGGCTTGGCTGAACAGTTCTGGAA	564
RESULT	9		
LOCUS	BP016149		
DEFINITION	BP016149	522 bp mRNA EST 10-OCT-2000	
VERSION	BP016149	uy1ae08.v1 NCI-CGAP.Lju30 Mus musculus cDNA clone IMAGE:3659630	5'
KEYWORDS	STRUTEN.	' similar to TR:O95793 O95793 DOUBLE-STRAND RNA-BINDING PROTEIN	
SOURCE	ORGANISM	house mouse.	
COMMENT	JOURNAL	Contact: Robert Strausberg, Ph.D.	
REFERENCE	AUTHORS	Email: Robert_Strausberg@nih.gov	
TITLE	TI	Tissue Procurement: Gilbert Smith, Ph.D.	
		cDNA library Preparation: Life Technologies, Inc.	
		DNA Sequencing by: Washington University Genome Sequencing Center	
		clone distribution: NCI-CGAP clone distribution information can be	
		found through the I.M.A.G.E. Consortium/LJNL at:	
		image.ljnl.gov/image/html/lresources.shtml	
FEATURES	SOURCE	MGI:1421958	
		Seq primer: -40RP from Gibco	
		High quality sequence stop: 441.	
		Location/Qualifiers	
		1..522	
		/organism="Mus musculus"	
		/strain="CZECH II"	
		/db_xref="taxon:10090"	



```

QY 2728 agctgacctgggaagagcatctgtgctccctagtcagccctg 2770
Db 480 AGCTGGCTGGGGAAGACATTGGCTCTCTAGTCCAGCCCTG 522

RESULT 11
LOCUS BE894295 752 bp mRNA EST 29-SEP-2000
DEFINITION 601437666F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3922529 5',
mRNA sequence.
ACCESSION BE894295
VERSION BE894295.1 GI:10356518
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 752)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC/DCFP/DRP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Inyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LCM744 row: d column: 18
High quality sequence stop: 654.
Location/Qualifiers
1. 752
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3922529"
/clone_lib="NIH MGC 72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT 223 a 194 c 192 g 143 t
ORIGIN
Query Match 17.8%; Score 508.4; DB 137; Length 752;
Best Local Similarity 84.8%; Pred. No. 4.1e-130;
Matches 616; Conservative 0; Mismatches 106; Indels 4; Gaps 4;

QY 892 attatgccaaggatgatacttattagtagactgtcagacatccagcaggaagaaag 951
Db 1 AATATGCCAGGAGATCATCGATTAGCCAGACTGCGCCAGATCCAGAGCAAGCAAAAAG 60

QY 952 agaagagccagatgactgtctctctacagaagagaggtctccacagctgcaggagattg 1011
Db 61 AGAAGGAGCCAGATACCGCTCTCCACAGAGCGGCGCTCCCGCGCGAGGAGTTTG 120

QY 1012 tgatgcaagtagaggttggtgcatcacactgcagaagaggtggtgtaacataagaagtgtg 1071
Db 121 TGATGTCAGGTGAAGTTTGAACACCACTGCGAAGAGAGAGGCGCACCAAGAAAGTTG 180

QY 1072 ccaagcgttaatctgtctgtagaacaatgctgagatcctggtgttcaagtctccagagcgc 1131
Db 181 CCAAGCGCAATGACAGCGAAGCAATGCTGAGATCCTGTTTCAAAATCCCGCAGCGCC 240

QY 1132 agctgtccaagccagcactcaatcagaagaagagactccagtaagaaccaggagagcg 1191
Db 241 AGCCACACCAACCGCACTCAAGTCAAGAGAGAGACACCAATTAAGAAACAGG-GATG 299

```

```

QY 1192 gaagaaagtaacgttttlltgaacctagccctggtggtatgaaatgaaactagtaacaag 1251
Db 300 GAAGAAAGTAACCTTTTGAACCTAGCTGCTGGGAGATGAATAATGGACTAGTATAAG 359

QY 1252 acgaggaattcagatgcttactatctacatcagcagctgcagctggaattctcccca 1311
Db 360 AGGATGAGTTCAGATGCTTATCTATCAAGTCAACACACTGCTCTGGAATTTCCCA 419

QY 1312 tgggtgccaagatttgccagagctgtcggtgttagtcaagagcaccacccaagaatttca 1371
Db 420 TGGTCCCGAGGTGCGCCAGGCTGTAGAGATTAGTCAAGACATCACACCAAGATTGTA 479

QY 1372 ccaagggcagctcccaaatctctcccaagcagcaagtaactgcatatgataccagagattgt 1431
Db 480 CCAGGACAGCTCCGATCTCCGATCTCCGACAGCGATACCTGATGATACCCGAGAGTTGT 539

QY 1432 tgaacggggagcactcgcgcacacagcgagacatlttaa-gagtaacatctctcaggc 1490
Db 540 AGTATGGGGGACCTCGCCACAGCCAGATCATTTTAACGAATACATCTCTTCAAGC 599

QY 1491 caggtaccccatggaacctgcgactagacctctgagcagctgactaccttccagagcc 1550
Db 600 CACGTACCCCATGGA-CTCTACAGAGATCTCTGAGCAA-TGGACTATCTTCCAGAGTC 657

QY 1551 caggtatccaggttgaatacaaaagatttcccaagaacaacaagagagtgatct 1610
Db 658 CAGGATTCGGGTTGATGCAAAAGACTTTCACCAAAAACACAGAGATTGTTCTCTATCA 717

QY 1611 ctcatc 1616
Db 718 ATGCTC 723

RESULT 12
AM542549/c 511 bp mRNA EST 31-AUG-2000
LOCUS C0158H10-3 Mouse E7.5 Extraembryonic Portion cDNA Library Mus
DEFINITION musculus cDNA clone C0158H10 3', mRNA sequence.
ACCESSION AM542549
VERSION AM542549.1 GI:7184966
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Scurionathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 511)
AUTHORS Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac
,M.J., Pantano,S., Sano,Y., Piao,Y., Nagataja,R., Dol,H., Wood,W.H.
III, Becker,K.G. and Ko,M.S.H.
Genome-wide expression profiling of mid-gestation placenta and
embryo using a 15,000 mouse developmental cDNA microarray
Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
20381348
COMMENT Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@nigmsun.grc.nia.nih.gov
Plate: C0158 row: H column: 10
Seq primer: -21M13 Forward
High quality sequence stop: 511
POLY-A-Yes.
Location/Qualifiers
1. 511
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="C0158H10"
/clone_lib="Mouse E7.5 Extraembryonic Portion cDNA
Library"
/sex="unknown"
/dev_stage="7.5dpc Embryo"
FEATURES
SOURCE

```



```

/lab_host="DH10B"
/note="Vector: pSPORT1 (Gibco/BRL Life Technology);
Site:1: SalI; Site:2: NotI. Total RNAs were extracted from
5 EPC. The double-stranded cDNA was synthesized by
Gibco's kit with an Oligo(dT) primer (NotI primer-adapter
from GibcoBRL)
(5'-pGACTAGTCTTATGATCGGAGCGGCCGCTTTTCTTTTCTTT-3')
from 0.8ug of mRNA. The double-stranded cDNAs were
treated with T4 DNA polymerase and purified by
ethanol-precipitation. The cDNAs were ligated to
lone-linker L1-SalI (include SalI sequence). The cDNAs
were purified by phenol/chloroform and separated from
free linkers by Centricon 100. Then, cDNAs were amplified
by long-range high fidelity PCR using Takara's Ex Taq
polymerase. Then, the cDNAs were purified by
phenol/chloroform and by Centricon 100. The cDNAs were
digested with SalI and NotI enzymes. Then, the cDNAs were
size selected by Gibco's Size Fractionation Column. The
cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid
vector. The DH10B E. coli host was transformed with the
ligation mixture by chemical method. The library was
constructed by Xiaohong Wang."

```

```

BASE COUNT      149 a      120 c      116 g      126 t
ORIGIN

Query Match
Best Local Similarity 99.6%; Pred. No. 5.3e-130;
Matches 509; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2345 ccggttggtggtgcatctctgctcgtgctggtgctgaacagttctggaattaccaccaga 2404
    |||||||
Db 511 cccggttggtggtgcatctctgctcgtgctggtgctgaacagttctggaattaccaccaga 452

QY 2405 atccctgactcctgcctcctgtataaattgacagcttagagactttaacttagatc 2464
    |||||||
Db 451 ATCCCTGACTCCCTGCCCTTGATTAATTTGACACCTTAGGACTTTTAACCTTTAGATC 392

QY 2465 aaaaagtatgctctttaaactttttaaggagcagacttaataatgaacccctgac 2524
    |||||||
Db 391 AAAAGATATGCTCTTTTAACTTATTTTAAAGAGACACTTTTAAAGAGACCCCTGAC 332

QY 2525 cttaccacattataacagaattgtcaaaaagagtgcttctttaaaggagtagcttttt 2584
    |||||||
Db 331 CTTTACCATTATAACAGAAATTTGTCAAAAAGAGCTTTCTTTAGAGAGAGCTTTT 272

QY 2585 taccacactacagagacattactcgttaggcccaagaactacagcgtggtgtccctagaag 2644
    |||||||
Db 271 TACCACTACAGAGGCGATTACTGTAGGCCCAAGAACTACAGCTGTGTCCCTAGAGG 212

QY 2645 gcccaataacagtaattccaaccttaagtgcgggaaaggtgacaggttccctggtgctg 2704
    |||||||
Db 211 GCCCAATACAGTCAATTCACAACTTAAGTCGGGAAAGGTGCGAGTTCCTGCTG 152

QY 2705 gtgtgacagagggcagagcagctgagcgtggtggaagaagcattgtgctctagtgca 2764
    |||||||
Db 151 GTGTGACAGAGGGCAGGACAGCTGAGCTGGCGGGAAGAGCAATGTGGCTCTAGTGCA 92

QY 2765 gccctgcttcacactctgtgttagctggaacctcccaactacatgaaataaatactc 2824
    |||||||
Db 91 GCCCTGCTTCACACTCTGTGTGTAGCTGGAACCTCCACACTCATGGAATATAGTAAACCTC 32

QY 2825 accttcttctgacccaataatgtataact 2855
    |||||||
Db 31 ACTTCTTTGTGCACCAATAATGTATTAATCT 1

RESULT 13
LOCUS BE674157 616 bp mRNA EST 08-SEP-2000
DEFINITION 7d/6a04.x1 NCI-CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3218862 3'
similar to TR:095793 095793 STRUFEN PROTEIN. ; mRNA sequence.
ACCESSION BE674157

```

```

VERSION BE674157.1 GI:10034698
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE 1 (bases 1 to 616)
JOURNAL NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq. primer: -400P from Gibco
High quality sequence stop: 469.
Location/Qualifiers
1. 616
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3218862"
/clone_lib="NCI-CGAP_Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: Lung; Vector: pT7AD-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI-CGAP_Lu24 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (clonoids
1414920-1417991 and 1520904-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo."

```

```

BASE COUNT      179 a      153 c      139 g      141 t      4 others
ORIGIN

Query Match
Best Local Similarity 89.0%; Pred. No. 2e-128;
Matches 540; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 1164 aagactccaaagaaacagagagacggaagaatgacgtttttgaactagacct 1223
    |||||||
Db 8 AAGCACCCCAATAAAGAACAGGGGATGAGAAAGAAATTAACCTTTTGAACCTGCTCT 67

QY 1224 ggggatgaaatggaactagtaacaagagcagagatgcagatgaccttaactagcat 1283
    |||||||
Db 68 GGGGATGAAATGGGATTAATTAATTAAGAGATGAGATTCAAGATGCTTATCTAAGTCAT 127

QY 1284 cagcagctgcagactggaattctcccaatggtgcggaagtgtcccaagctgtgcgggtc 1343
    |||||||
Db 128 CAGCAGCTGCTGCTGTGAATCTTCCCATGTGCTCCAGAGTGTGCGCCAGCTGTGAGGAGTT 187

QY 1344 agtcaagacacacacacacaaagatttcacacagggcagctcacaatctctgccaagaagac 1403
    |||||||
Db 188 AGTCAGAGACATCACACAAAGATTTTACAGAGGCACTCCGAATCTCTGCCAAGGCCACG 247

QY 1404 gtaactgcatagtacccagagatgtgtgtaaggggacactgcccacagccgaagacc 1463
    |||||||
Db 248 GTAACGTCCATGATAGCCGAGAGATTGTGTATGGGGGCACTCGCCACAGCCGAGACCC 307

QY 1464 attttaagagataacatctcttaagcagcagttacccatgagactcgtgactagaacctct 1523
    |||||||
Db 308 ATTTTAAAGATAACATCTCTTCAGGCCACGTACCCCATGAGACTCTCACAGAGACCTCT 367

```



Email: cdna@igsun.grc.nia.nih.gov  
 Plate: C0146 row: H column: 10  
 Seq primer: 21M13 Forward  
 High quality sequence stop: 497  
 POLY(A)-Yes.

## FEATURES

## SOURCE

## Location/Qualifiers

1. 497  
 /organism="Mus musculus"  
 /strain="C57BL/6j"  
 /db\_xref="taxon:10090"  
 /clone="C0146H10"  
 /clone\_lib="Mouse E7.5 Extramembronic Portion cDNA  
 library"

/sex="unknown"  
 /dev\_stage="7.5dpc Embryo"

/lab\_host="DH10B"

/note="vector: pSPORT1 (Gibco/BRL Life Technology);  
 site\_1: SalI; site\_2: NotI; Total RNAs were extracted from  
 5 EPC. The double-stranded cDNA was synthesized by  
 gibco's kit with an Oligo(dT) primer (NotI primer-adaptor  
 from GibcoBRL)

[5'-GACTAGTTCATGATCCGAGCGGCCCTTTT-3']  
 from 0.8ug of mRNA. The double-stranded cDNAs were  
 treated with 74 DNA polymerase and purified by  
 ethanol-precipitation. The cDNAs were ligated to  
 lone-linker IL-Sal3 (include SalI sequence). The cDNAs  
 were purified by phenol/chloroform and separated from  
 free linkers by Centricon 100. Then, cDNAs were amplified  
 by long-range high fidelity PCR using Takara's Ex Taq  
 polymerase. Then, the cDNAs were purified by  
 phenol/chloroform and by Centricon 100. The cDNAs were  
 digested with SalI and NotI enzymes. Then, the cDNAs were  
 size selected by Gibco's Size Fractionation Column. The  
 cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid  
 vector. The DH10B E. coli host was transformed with the  
 ligation mixture by chemical method. The library was  
 constructed by Xiaohong Wang."

BASE COUNT 145 a 111 c 113 g 128 t  
 ORIGIN

Query Match 17.0%; Score 486; DB 92; Length 497;  
 Best Local Similarity 99.8%; Pred. No. 6.1e-124;

Matches 497; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 2358 attctgtgctctggtgcttgaagattctgaattaccaccagaatcctgactccc 2417  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 497 ATTCTGTGCTCTGGCTTGAGCTGACAGTTCTGGAATTACACAGATCTTGACTCCC 438  
 QY 2418 tgcacctgtataaattgacacagcttagaactttaacttagatcaaaagaatgctc 2477  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 437 TGCCCTTGATTAATTTGACACACTTAGACCTTTAACTTAGATCAAAAGATATGCTC 378  
 QY 2478 ctlttaacttaattttaagagacgaactttaaatgagccctgaaccttaccattat 2537  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 377 CTTTAACTTAATTTTAAAGAGACACTTTAAATGAGCCCTGACCTTACCATTAAT 318  
 QY 2538 aacagaatttgcaaaagagtggttcttgagaggtagctttttttaccacatacag 2597  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 317 AACAGAAATTTGTCAAAGAGAGTGTCTTGAGAGGATAGC-TTTTTTTACACACTACAG 259  
 QY 2598 gacataacctgtagcccaagaagactacagagctggtctccctagagagcccaatacagtc 2657  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 258 GACATTACTGTAGGCGCCAGAAACACTACAGAGCTGGTGTCCCTAGAGGCGCCCAATACAGTC 199  
 QY 2658 aattcaacctctaagtcgagggaaggtgacaggttccctgctgctggtgtgacagggg 2717  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 198 AATTCCAACCTCTAACTCGGGGAAAGGTGACAGAGTTTCTGAGTGTGTGCACAGGGG 139  
 QY 2718 caggcaggtcaggtcgtggggaagaagatctgtgctcctagtgcagccctcttccac 2777  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 138 CAGGCAAGTCACTGGCTGGGGAAGAGCATGTGTGCTCCTAGTGACAGCCCTGCTTCAC 79

QY 2778 tcttggttagctggaaccttcccactcalggaatataaagtaaacctacttcttgcga 2837  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 78 TCTTGGTTAGCTGGAACCTTCCACATGGAATFAAAGTAACACACTTCTTGTCA 19  
 QY 2838 ccaataatggttaact 2855  
 ||||||||||||||||||||  
 Db 18 CCAATAAATGTAACTACT 1

Search completed: April 5, 2001, 00:30:47  
 Job time: 53143 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 4, 2001, 09:45:04 ; Search time 11371 Seconds  
(Without alignments)  
1936.289 Million cell updates/sec

Title: US-09-316-048-1  
Perfect score: 3142  
Sequence: 1 acttcctgcgcggctcgcgg.....taatactaaaaa 3142

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues  
Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: em\_estba:\*  
42: em\_estfun:\*  
43: em\_esthum1:\*

44: em\_esthum2:\*  
45: em\_esthum3:\*  
46: em\_esthum4:\*  
47: em\_esthum5:\*  
48: em\_esthum6:\*  
49: em\_esthum7:\*  
50: em\_esthum8:\*  
51: em\_esthum9:\*  
52: em\_esthum10:\*  
53: em\_esthum11:\*  
54: em\_esthum12:\*  
55: em\_esthum13:\*  
56: em\_esthum14:\*  
57: em\_esthum15:\*  
58: em\_esthum16:\*  
59: em\_esthum17:\*  
60: em\_esthum18:\*  
61: em\_esthum19:\*  
62: em\_esthum20:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estov1:\*  
68: em\_estov2:\*  
69: em\_estp11:\*  
70: em\_estp12:\*  
71: em\_estp13:\*  
72: em\_estp14:\*  
73: em\_estp15:\*  
74: em\_estro1:\*  
75: em\_estro2:\*  
76: em\_estro3:\*  
77: em\_estro4:\*  
78: em\_estro5:\*  
79: em\_estro6:\*  
80: em\_estro7:\*  
81: em\_estro8:\*  
82: em\_estro9:\*  
83: em\_estro10:\*  
84: em\_estro11:\*  
85: em\_estro12:\*  
86: em\_estro13:\*  
87: gb\_est41:\*  
88: gb\_est42:\*  
89: gb\_est43:\*  
90: gb\_est44:\*  
91: gb\_est45:\*  
92: gb\_est46:\*  
93: gb\_est47:\*  
94: gb\_est48:\*  
95: gb\_est49:\*  
96: gb\_est50:\*  
97: gb\_est51:\*  
98: gb\_est52:\*  
99: gb\_est53:\*  
100: gb\_est54:\*  
101: gb\_est55:\*  
102: gb\_est56:\*  
103: gb\_est57:\*  
104: gb\_est67:\*  
105: gb\_est68:\*  
106: gb\_est69:\*  
107: gb\_est70:\*  
108: gb\_est71:\*  
109: gb\_est72:\*  
110: gb\_est73:\*  
111: gb\_est74:\*  
112: em\_esthum21:\*  
113: em\_esthum22:\*  
114: em\_esthum23:\*  
115: em\_estom1:\*  
116: em\_estom2:\*

117: em\_estp16:\*  
 118: em\_estp17:\*  
 119: em\_estp18:\*  
 120: em\_estp19:\*  
 121: em\_estp20:\*  
 122: em\_estp21:\*  
 123: em\_estp22:\*  
 124: em\_estp23:\*  
 125: em\_estp24:\*  
 126: em\_estp25:\*  
 127: em\_estp26:\*  
 128: em\_estp27:\*  
 129: em\_estp28:\*  
 130: em\_estp29:\*  
 131: em\_estp30:\*  
 132: em\_estp31:\*  
 133: em\_estp32:\*  
 134: em\_estp33:\*  
 135: em\_estp34:\*  
 136: em\_estp35:\*  
 137: em\_estp36:\*  
 138: em\_estp37:\*  
 139: em\_estp38:\*  
 140: em\_estp39:\*  
 141: em\_estp40:\*  
 142: em\_estp41:\*  
 143: em\_estp42:\*  
 144: em\_estp43:\*  
 145: em\_estp44:\*  
 146: em\_estp45:\*  
 147: em\_estp46:\*  
 148: em\_estp47:\*  
 149: em\_estp48:\*  
 150: em\_estp49:\*  
 151: em\_estp50:\*  
 152: em\_estp51:\*  
 153: em\_estp52:\*  
 154: em\_estp53:\*  
 155: em\_estp54:\*  
 156: em\_estp55:\*  
 157: em\_estp56:\*  
 158: em\_estp57:\*  
 159: em\_estp58:\*  
 160: em\_estp59:\*  
 161: em\_estp60:\*  
 162: em\_estp61:\*  
 163: em\_estp62:\*  
 164: em\_estp63:\*  
 165: em\_estp64:\*  
 166: em\_estp65:\*  
 167: em\_estp66:\*  
 168: em\_estp67:\*  
 169: em\_estp68:\*  
 170: em\_estp69:\*  
 171: em\_estp70:\*  
 172: em\_estp71:\*  
 173: em\_estp72:\*  
 174: em\_estp73:\*  
 175: em\_estp74:\*  
 176: em\_estp75:\*  
 177: em\_estp76:\*  
 178: em\_estp77:\*  
 179: em\_estp78:\*  
 180: em\_estp79:\*  
 181: em\_estp80:\*  
 182: em\_estp81:\*  
 183: em\_estp82:\*  
 184: em\_estp83:\*  
 185: em\_estp84:\*  
 186: em\_estp85:\*  
 187: em\_estp86:\*  
 188: em\_estp87:\*  
 189: em\_estp88:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1757.8	55.9	1801	146	U69197 U69197 Soar
2	701	22.3	1018	136	BE867390 601442470
3	646.2	20.6	1036	137	BE889880 601512132
4	621.2	19.8	724	97	AW952360 EST364430
5	617.2	19.6	752	137	BE894295 601437666
6	606.4	19.3	821	107	BE379687 601159257
7	604	19.2	609	107	BE380123 601159257
8	603.8	19.2	686	110	BE613972 601504093
9	602.2	19.2	674	38	AV707406 AV707406
10	596.4	19.0	829	106	BE300155 600944602
11	594.4	18.9	616	111	BE674157 7476a04.x
12	588.2	18.7	594	40	AM162002 av72c03.x
13	571.4	18.2	612	38	AV695897 AV695897
14	565.4	18.0	749	110	BE617223 601441906
15	555	17.7	609	136	BE810330 MR4-PT005
16	551.6	17.6	615	134	BE082712 RC2-BT064
17	550	17.5	550	107	BE348298 hw16b01.x
18	548	17.4	564	40	AM163206 au97f05.x
19	542.4	17.3	584	136	BE813737 RC1-BN003
20	534	17.0	673	135	BE781223 601469159
21	531.2	16.9	536	136	BE837505 RC2-FN009
22	531	16.9	866	136	BE868328 601443887
23	530.2	16.9	573	3	AA206573 zq51f08.r
24	524.8	16.7	831	106	BE299828 600944602
25	517.8	16.5	910	138	BE966164 601660093
26	513	16.3	543	134	BE089861 RC5-BT070
27	511.4	16.3	513	13	AA889669 ak5a04.s
28	510.4	16.2	512	27	A1983007 w146g02.x
29	508.8	16.2	852	137	BE872137 601446238
30	500.6	15.9	673	109	BE541462 601067913
31	497.4	15.8	633	89	AW320695 uc02b10.y
32	488.2	15.5	569	3	AA191622 zp81g10.s
33	487	15.5	516	25	A1819766 w142a10.x
34	483.4	15.4	485	25	A1802592 w116a03.x
35	478.4	15.2	480	134	BE042598 hc26e03.x
36	475	15.1	494	89	AW351909 RC0-CT020
37	470.4	15.0	509	142	N31181 N31181
38	470	15.0	517	7	AA427366 zw33c05.r1
39	470	15.0	517	92	AA580332 IL2-HFN04
40	467.4	14.9	703	137	BE888420 601518177
41	465.6	14.8	686	40	AW160378 au72c03.y
42	457	14.5	724	38	AV700494 AV700494
43	455.2	14.5	561	141	H16250 ym13d10.s1
44	451.8	14.4	455	96	AW874257 hp96g04.x
45	450.4	14.3	452	17	AT190464 qd38b02.x

## ALIGNMENTS

RESULT 1  
 U69197 1801 bp mRNA EST 27-OCT-1999  
 LOCUS U69197 Soares infant brain INIB Homo sapiens cDNA clone 22368, mRNA  
 DEFINITION sequence.  
 U69197  
 ACCESSION U69197.1 GI:2739420  
 VERSION U69197.1  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 1801)  
AUTHORS Volotio,S., Simon,G., Repetto,M., Cucchiari,M., Banfi,S., Borsani  
G., Ballabio,A. and Zollo,M.  
TITLE Sequencing analysis of forty-eight human image cDNA clones similar  
to drosophila mutant protein  
JOURNAL DNA Seq. 9 (5-6), 307-315 (1998)  
MEDLINE 99452388  
COMMENT Contact: Zollo, Massimo  
Teleton Institute of Genetics and Medicine  
Via Olgettina 58, Milan, MI 20132, Italy  
Email: zollo@tigem.it.  
LOCATION/Qualifiers  
1. 1801  
FEATURES  
source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="19p13.1"  
/clone="22368"  
/sex="female"  
/clone.lib="Soares infant brain INIB"  
/dev\_stage="73 days post natal"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: whole brain; Vector: laimid BA; Site:1: Not  
I; Site:2: Hind III; 1st strand cDNA was primed with a Not  
I - oligo(dT) primer [5].  
ACTGGAAGAATTCGGCGCCGACAGAAATTTTATTTTATTTT 3';  
double-stranded cDNA was ligated to Hind III adaptors  
(Pharmacia), digested with Not I and directionally cloned  
into the Not I and Hind III sites of the laimid BA vector.  
Library went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 503 a 409 c 386 g 503 t  
ORIGIN

Query Match 55.9%; Score 1757.8; DB 146; Length 1801;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1770; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1371 gctccgaatcctgcgaagcagcagtaactgcatactagatccgagatgtgtatgag 1430  
|||||  
Db 5 GCTCGAATCCTGGCAAGGCGTACCTGCATGATAGCCCGAGATTGTGTATGGG 64  
|||||

QY 1431 ggcacctcgcgcacagcgcagacatttaagaataacatccttccagccagctacc 1490  
|||||  
Db 65 GGCACCTCGCCACAGCCGAGACCATTTTAAAGATAACATCTCTTCAGGCCACGTACC 124  
|||||

QY 1491 catgacctcaagagagccctcaggaactgactatcttccagatccagagatc 1550  
|||||  
Db 125 CATGACCTCTACGACGACCTCTGAGCACTGACTATCTTTCCAGAGTCCAGGGATTTC 184  
|||||

QY 1551 caggctgaatacaagactccccaacaacaagaagacgaattgtatctcttcaat 1610  
|||||  
Db 185 CAGGTTGAATCAAAAGACTTCCCAAAAACAAGAAGCAATTGTCTCTTATCAT 244  
|||||

QY 1611 tgcctcctcagcaacctctgatacgcaatgttatcggcaagatgtgagtcctgcat 1670  
|||||  
Db 245 TGCTCTCTTCAGCCACCTCTGATCAGCCATGTATCGGCAAGATGTGAGTCTCGCAT 304  
|||||

QY 1671 gatattgctgcgtgaataattaaagtgtctgtcgtgagttggaaccaaaagtaacag 1730  
|||||  
Db 305 GATATGGCTGGCTGTAACATCTTAAAGTTGCTGCTGAGTTGGACCAACAAAGTACAGAG 364  
|||||

QY 1731 atgccaaagaagaagaagcaaatgtctgtgtgtgagagtgctgaacctttctggc 1790  
|||||  
Db 365 ATGCCAAGAACAGGAAGAGCAATGTCTGTGTGGAGGTGCTGAACCTTTCTGGC 424  
|||||

QY 1791 catgacacataataatcccaacataatactgaataaactgtaactgtcttgaat 1850  
|||||  
Db 425 CATGAACCATTAATAATCCCAACATATATACGTAAATACTGAAACTGCTTGAATAAT 484  
|||||

QY 1851 tgaattcttgatactccctcagctgagccgagagacaggtgtgtaagaatgtggtcagca 1910  
|||||  
Db 485 TGAATTTCTGATACCTTCACATGGGCGAGACACGCTGGGTAAAGATGTGGGACAGA 544  
|||||

QY 1911 gcaagggaagaacaagaacaagaagagcgctgtggcc-ggcttgactgtgtctggagt 1969  
|||||  
Db 545 GCAGGGAAGACAAACAGAAACAAAGAGCGGCTGTGGCCCGGGTGGCTGTGTGGGT 604  
|||||

QY 1970 ttgtgtgataagcacttggtgagacttgagcgtgtcccttaagcaatgcaactgctgtgagg 2029  
|||||  
Db 605 TTGTTTGATGATGCGCACTGGTGACTGGCGGTCTCCCTACGCAATATGCACTCTGTGGG 664  
|||||

QY 2030 aagaagggtctgcccagcagctggttctcccgaggacacagatccacaccttgagca 2089  
|||||  
Db 665 AAGAAAGGCTGCCAGCAGCTGTGTCCTCCGGGACACACAGATCCACACCTGGGCA 724  
|||||

QY 2090 cctccgtgttggctcttttttccctgtgtgaaagaagaagcgcaagccctctc 2149  
|||||  
Db 725 CCTCGGTGTGTGCTTTTTCCTTCCTGTGAAAGAAAGGAGCGACGACCCCTTCTC 784  
|||||

QY 2150 aagcttgatcactcagacacatlggagacaacccctggacagccatgccaagagagcct 2209  
|||||  
Db 785 AAGCTGCTCAGTACAGACATGAGACAAACCTGGACAGCCATGCCAGAGAGGCTT 844  
|||||

QY 2210 tgaacggcccccagagactaaagacacagagaaatacaatgcttccctactcaagctgac 2269  
|||||  
Db 845 TTGAACGGCCCCAGAGCTAAAGACACGAGAAATATCAATATGCTTCTCAGAGGTAC 904  
|||||

QY 2270 ccaacttcttagtgtgcaagcgcgcacacacacccctcctcgtcagtaaccacatacgaact 2329  
|||||  
Db 905 CCAACTTTCTAGTGTGCGACAGGCCCCACACCTCTCTGACATCCACCATCAGCACT 964  
|||||

QY 2330 gcttctctcacaagagactgtatctatgttaatttaatttcttcttgaatgata 2389  
|||||  
Db 965 GCTTCTCTCAACAGTATGATCTGATTTAGTTTCATTAATTTCTTATGATGATATG 1024  
|||||

QY 2390 acaactataaatttcatcttgagaatttctcaatctgactgatactgataaagaacagt 2449  
|||||  
Db 1025 ACACATATTAATAATTTTATTTGACAAATTTCTCAATTTGATCTAOTTAATAAGACATTT 1084  
|||||

QY 2450 tggaaacttgctgagactgactatcaataatcaacgacaagaalcatlcatgt 2509  
|||||  
Db 1085 TGGAAACTTGCTGAGACTGACTTATATCAATATATCAACGACAAAGATCATATTCATGT 1144  
|||||

QY 2510 gtaatgtgtgaacatttatttcaatgtactaaaccccgagagcttgaagatgacaat 2569  
|||||  
Db 1145 GTATGTGTTGAGACATTTTATTTTCAATGACTTAACCCGAGCAGTTTTCAGTATGCAAT 1204  
|||||

QY 2570 tgtgtgccctcgtgtgagcgtgaacagctcctggaacttcaaaaaccttgaaatgctc 2629  
|||||  
Db 1205 TGTGTGCCCTTGTTGCTGACGTGAAACAGTCTCGGACTTTCAAAACCTTGAATTAAGTCTC 1264  
|||||

QY 2630 ccacagctgtataaatttgaaatattgaatatttaaccttgaatgatacatlctgtctc 2689  
|||||  
Db 1265 CCACAGTTGTAAATTTGACAAATTTAGAAATTTAACTTTAGATGATCAATTTGGTCTC 1324  
|||||

QY 2680 atttatttcaatttatttlttltgtaatgcaaaagaaacttaaatgaaacttgatcct 2749  
|||||  
Db 1335 ATTTTATTTTCAATTTTATTTTGTAAATGCAAAAGAGACTTAAATGAACTTTGATCTCT 1384  
|||||

QY 2750 gtttaagaatctataaacaacatgtglatcatalacataagccttgagaaacttaact 2809  
|||||  
Db 1385 GTTTTAAAGATTTATTAATAAACATTTGTATCTATACATATAGCGCTTGAGAGACTTACT 1444  
|||||

QY 2810 ttcactacactaagaalatalgatctcaatgtatgcatatataaocctgcaagatgatttc 2869  
|||||  
Db 1445 TTCACTACACTACAGAGATATATCTCATGTATGATCAATTAACCTGCAGAGTGAATTTTC 1504  
|||||

QY 2870 cagagtgctgactactgtttaaataatcattcaatcattaggcttgaagaagaaatgacatg 2929  
|||||  
Db 1505 CAGAGTGTCTGATCTGTTAATTTACATCTCAATTAAGGCTGAAAGAAATGACATACGTTT 1564  
|||||

QY 2930 ctgtatacagctgtgtccttctgtgtgttactgttacaagaagtgctgtgactga 2989  
|||||

|||||  
Db 1565 CTGTATACAGCTGTGCTTTGATGTTGTACTGATACAGACGAAGTGTGCACTGA 1624  
QY 2990 ggctctggcgtgtgtgtcgtatagaaacccgtgagccctgcgagtgtaagtaacgtctcca 3049  
Db 1625 GGCTGTGCGGTGTGCTGATGAAAGAGCTGTGAGCCCTGCGAGTTAAGTACGCTTCCA 1684  
QY 3050 ttcatgtttacgctggaatttctcccatggaatgtaagaactaaagtgtttgt 3109  
Db 1685 TTCAATGTTTACGCTGGAATTTTCTCCCATGTGAATGTAAAGTAACTAATGTTTGT 1744  
QY 3110 catcaataatggttaactactaaataaaaaa 3142  
Db 1745 CATCAATTAATGCTAATACTAAAAAAA 1777  
RESULT 2  
BE867390 1018 bp mRNA EST 27-SEP-2000  
LOCUS 601442470F1 NIH\_MGC\_65 Homo sapiens cDNA clone IMAGE:3846689 5',  
DEFINITION mRNA sequence.  
ACCESSION BE867390  
VERSION BE867390.1 GI:10316166  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 1018)  
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM546 row: 1 column: 18  
High quality sequence stop: 654.  
Location/Qualifiers  
1. 1018  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3846689"  
/clone\_1ib="NIH\_MGC\_65"  
/tissue\_type="adenocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: Colon; Vector: pCMV-SPOK6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.8 kb. Library constructed by Life  
Technologies."   
BASE COUNT 295 a 267 c 262 g 194 t  
ORIGIN  
Query Match 22.3%; Score 701; DB 136; Length 1018;  
Best Local Similarity 90.0%; Pred. No. 8.9e-162;  
Matches 809; Conservative 0; Mismatches 80; Indels 10; Gaps 5;  
QY 916 tggccagatcagcagcagcaaaagagaagaccagatcacgctctccacagagc 975  
Db 1 TGCCCCAGATCCAGCAGCAAAAAGAGAGAGCAGACGCTCCCTCAGACAGC 60  
QY 976 gaggctcccgcgcgagggaggttgatgtagtgtaaggttggaacacacacgcag 1035  
Db 61 GAGGCTCCCGCGCCAGAGAGAGTGTGTGATGAGGTTGGAACACACACTGCAG 120  
QY 1036 aaggacgagccacacaaagaagtgccaagcgaatgcagcgagaacatgctggaga 1095

|||||  
Db 121 AAGGAACGGCGACCAACAAGAGGTGGCCAAAGCGCAATGCAGCCGAGACATGCTGTGAGA 180  
QY 1096 tcttggtttcaaaagtcctccgcaagcgagccacccaacccgacacaaatcgaagagaga 1155  
Db 181 TCTTGGTTTCAAAAGTCCCGAGCCGAGCCACCAAAACCCGACATCAATCTAGAGAGA 240  
QY 1156 agacaccataaagaacccagggagatggaagaagttaaccttttgaacctgacctg 1215  
Db 241 AGACACCCATTAAGAAACAGAGGAGATGAGAAAGTAACCTTTTGAACCTGGCTCTG 300  
QY 1216 gggatgaaatgaggactagtaataaagaga tgaagtcgaagtccttatcgaatc 1275  
Db 301 GGGATGAAATGGGACTAGTAATAAGAGATGAGTTCAGGATGCTTATCTAAGTCATC 360  
QY 1276 agcagctgctgttggaatcttcctccatggtgcccagagtcgcccagagctgtagagat 1335  
Db 361 AGCAGCTGCTGCTGGAAATCTTCCATGATGTCGCGAGGTGCGCAGGCTGTAGAGTTA 420  
QY 1336 gtcaaggacacacacacaaagatttaccagaggcagctccgaaatcctgccaaggcagcg 1395  
Db 421 GTCAAGGACATCACACCAAGATTTTACAGGGCAGCTCCGAATCGTCCAGAGCCACGG 480  
QY 1396 taactgccatgataagccgagagtggtgta tgggggaccttgcccaagccgagacca 1455  
Db 481 TAACTGCATGATAGCCCGAGAGTTGCTATGGGGCACCTCGCCACAGCCAGACCA 540  
QY 1456 tttaagaataacatctcttcagccaggtaccatgagactctcagaagaccctctg 1515  
Db 541 TTTTAAAGATAACATCTCTTCAGGCGACGTACCCATGAGCTCTCAGAGACCTCTG 600  
QY 1516 agcaactgagactctcttcagagtcgaagtcaggttccaggttgaatacaaaagctccca 1575  
Db 601 AGCACTGAGACTATCTTTCAGAGTCCAGGATTCAGGTGAATGAAGATTTCCCA 660  
QY 1576 aaaaacaagaagaatttgatct---ctatcaatgtctctctcagcaacctgtga 1632  
Db 661 AAATATCACCAAGAACCAATTCGCTATCTCTTATCAATGCTCTCTCACCCACTGTGAT 720  
QY 1633 tcaagcatgtatcggcaagaagatgtgagtcctgca---tgatagtcctgctgtaaca 1689  
Db 721 CCACATGATGATTCGGAGGTTGTGGGAATCTCGCAATGATCATGGTGGCTGGACA 780  
QY 1690 tcttaaaag-ttgcgtgctgagttggaaccaaaa-gtacagagatgccaagaacagga 1747  
Db 781 TCTTTAAGATGCTGCGGAGAGTGAAGCCCAAAAGTCAAAAATTTCCAGAAACGGGA 840  
QY 1748 cggaccaaagtcgtgtgtgtggg--aggctgaaccttctctgccaatgacatata 1804  
Db 841 GGGACCCAGTTCGTTGTGCGACAGGTCTGAAACATTTTGGGCTGAACCATTA 899  
RESULT 3  
BE889880 1036 bp mRNA EST 29-SEP-2000  
LOCUS 601512132F1 NIH\_MGC\_71 Homo sapiens cDNA clone IMAGE:3913590 5',  
DEFINITION mRNA sequence.  
ACCESSION BE889880  
VERSION BE889880.1 GI:10347645  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 1036)  
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.









QY	2309	agttaccacacacatcacacacacggtcttcctcccaacagtgatcgtatctcttaattcat	2368
Db	121	AGTACCACACCATCATCCACGTCTTCTCTTCCACACATGATCTGTATCTTTAGTTTCAT	180
QY	2369	tattctctcttgattgatacatgacactatataaaatlctcatcttgagaattctcaatgct	2428
Db	181	TATTTTCTTTTGATTGATGATATGACACTATATATAAATTTTCATTGAGAAATTTCTCAATTTGT	240
QY	2429	atctagttaaatagcacagtttggaactgtctgagacgtacttlatcaataactaac	2488
Db	241	ATCTGATTAAATAGCACAGTTTGGAAACTGCTGCAGACTGACTTATCAATAATCTAAC	300
QY	2489	cgacaagaatcaatcacatctgtatggttgtaagacattttatctatcttgacacccag	2548
Db	301	CGACAAAGATCATATCCATGATGTATGTGTAGACATTTTATTTTCATGACTAACCCAG	360
QY	2549	gacagtttcagtgatgacaaatctgtgtgcacctgtgtcagctgaacacagctctgacctt	2608
Db	361	GACAGTTTCAGTGAATGCAAAATTTGTGTGCCCTCTGTTCCAGCTGAACAGTCTGTGACTTT	420
QY	2609	caaaaaccttgaataagctctccacagctgtataaatlgacaalttgaagattttaaac	2668
Db	421	CAAAAACCTTGATATAGCTCCACAGAGTTGTATTAATTTGACAAATTTAGCAATTTTAAAC	480
QY	2669	tttagatgatacttggtgtgcattttttatctcatctttatcttggtaagtcaaacagga	2728
Db	481	TTTAGATATATCATTTTGGTTCATTTTTTATTTTCATTTTTTATTTTGTATATGCAACAGGA	540
QY	2729	cttaaatgaacttgcctcgttctttaaagaattlaaaaaacactgtgatactacat	2788
Db	541	CTTAATGACATTGTGATCTGTGTTTAAAGATTTATTAATAAACAATTTGATATCAT	600
QY	2789	atggg 2792	
Db	601	ATGG 604	
RESULT	8		
LOCUS	BE613972		
DEFINITION	601504039F1 NIH_MGC_71 Homo sapiens cDNA IMAGE:3905783 5',		
ACCESSION	BE613972		
VERSION	BE613972.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 686)		
JOURNAL	NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Tel: (301) 496-1550		
	Email: Robert.Strausberg@nih.gov		
	Tissue Procurement: ATCC		
	CDNA Library Preparation: Life Technologies, Inc.		
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNL at:		
	http://image.lnl.gov		
	Plate: LCM700 row: j column: 24		
	High quality sequence stop: 675.		
FEATURES	Location/Qualifiers		
source	1..686		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:3905783"		
	/clone_id="NIH_MGC_71"		
	/tissue_type="telomysarcoma"		

Query Match	19.2%	Score 603.8;	DB 110;	Length 686;
Best Local Similarity	97.1%	Pred. No. 76-138;		
Matches 669;	Conservative	0;	Mismatches 12;	Indels 8; Gaps 5
BASE COUNT	226 a	156 c	196 g	108 t
ORIGIN				
	/lab host="DH10B (phage-resistant)"			
	/site="Organ: uretus; Vector: pcmV-SpORt6; Site_1: NotI;			
	Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT			
	Average insert size 2.1 kb.			
QY	430	aacagatgctgcctgcacaaagcgcttgagatctctgagaatgagccctgcacagagagc	549	
DB	1	AACACGATGCTGCTGCCAAAGCGTTGAGGAYCCTGGAGATGAGCCCTGCCAGAGAGCG	60	
QY	550	tggaggtgaatgagagagaaatccgaagaanaaatctcataatctgaataagtaag	609	
DB	61	TGGAGGTGAATGGAAGAGAAATCCGACAGAAAATCTCATATAATCTGAATAAGTCAAG	120	
QY	610	tgcttgagatgtcacctiaaacygaacttgctgtgaatttcgaggltgagccyggagatg	669	
DB	121	TGTTTGAGATGTGCACCTTAAAGGAGACTTGCCTGTGAATTTTCGAGGGGCGCCGGAGAGTG	180	
QY	670	gccaccccacatagaagaacttgctgacaaaggttccgttgaggagttgtggggaag	729	
DB	181	GCCACCCACCATATMAAGAAC-TGTGTGCCAAGGTTTCGGT--GGGAGTTGTGGGGGAAG	237	
QY	730	gtgaagggaagaagaagaatcttcaagaanaaatgacgcaatgctgtcttgaggagc	789	
DB	238	GTGAAGGGAAAAGCAAGAGATTTCAAAGAAAATCCGCCATAGCTGTCTTTTGAGGAGC	297	
QY	790	tgaagaagtlacccgcctctgcctgcagtlgaacgaatgaagcctagaatcaaaagaanaa	849	
DB	298	TGAAGAAGTTACCGCCCTCGCTGCGATGTAACGAGTAAGACTGAATCAAAAAGAAAA	357	
QY	850	caaaacccatagtlcaagccacagacaagccagaaatatatgccagaggatcaatccgat	909	
DB	358	CAAAAACCATAGTCAACCCACAGACCAAGCCACCAATATGAGCCA--GGGATCAATCCGATT	416	
QY	910	ggcgacttggcccagaatccagcagagcaaaaagagagaagagccagagtaacagctcc	969	
DB	417	GCCGACTGCGCCCAATCCACAGAGCAGAAAAGAGAGAGACGACAGTACACGCTCTCTCA	476	
QY	970	caagagagagcctcccgagccgcagcagagagtttgaatgcaggttgaaggttcgaaac	1029	
DB	477	CAGAGCGAGCGCTCCGCGCGCGCAGGAGGAGTTGTGATGACAGTGAAGGTTGGAAMAC	536	
QY	1030	ctcgagaagaagagcgagccacaacagaagaagltggtccaaagcgcaatgcagccgaaga	1089	
DB	537	CTCGAGAAGAGAACGGGCACCAACAAAGAGTGCGCAAGCGCATCGAGCGAATCATGC	596	
QY	1090	tggagatctccttgg--tttcaaatgcccagcagcgagacccacaacaccgcgactaaag	1148	
DB	597	TGGAGATCTCTTGTTTCAAGTCCGCGAGGCGCAGCACCAACAA--CCGACTCAAGTCA	653	
QY	1149	gaagagaagaacacccataaagaacacagg	1177	
DB	654	GAGGAGAAGACACCCATAAAGAAACAGGG	682	
RESULT	9			
AV707406	674 bp	MRNA	EST	09-OCT-2000
LOCUS	AV707406	ADB Homo sapiens	CDNA clone	ABDCM03 5', mRNA sequence.
DEFINITION	AV707406	ADB Homo sapiens	CDNA clone	ABDCM03 5', mRNA sequence.
ACCESSION	AV707406	1	GI:10724671	
VERSION	AV707406.1	GI:10724671		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
REFERENCE	1 (bases 1 to 674)			



```

OY 258 acccctactgtaaacataatgcatctgcatgaaacttggaataaaacccaatgataag 317
    |||||||
Db 564 ACCCCCTACTGTAGACTAATATGCACTGTGATAACTTGAAAAAACCAATGTATAAG 505
OY 318 cctgttgacctactctcggatgagtcacactataacatacagatgaggaagtgct 377
    |||||||
Db 504 CCGTGTGACCTTACTCTCGATGATGATCCACTTAACTAACATGAGAGAGGTGCT 445
OY 378 tatccccgaggaagtaattacccattcccaagtcacacttactatataagtgaaact 437
    |||||||
Db 444 TATCCCCGAGGAGTACTTTTACCCATTTTCCAGTTTCACTTTTATCAAGTGAACCTT 385
OY 438 tctgttgaggaagcaagcaatttaattgcaagaagaacaaagacagagctgcgaacacat 497
    |||||||
Db 384 TCTGTGGAGGAGCAAGCAATTTATGCGAAAGAAAGACAGACAGCTGCGAAACAGAT 325
OY 498 gctgtgtgccaagaagcgttgagatcctgcagaatgagccctgcagaagagctgagtg 557
    |||||||
Db 324 GCTGTGCGCAAGCGTTGAGGATCTGCAGAAATGAGCCCTGCGCAGAGAGCGCTGAGGTG 265
OY 558 aatggaagaagaatcgaagaagaataatcctaataatctgaataatgaagtgattgag 617
    |||||||
Db 264 AATGGAAGAAGAAATCCAGAGAGAAATCTCAATAAATCTGAATTAAGTCAAGTGTGAG 205
OY 618 atgcacttaaacggaagctgctgtgaaatttcgaagtgagccggagagatgagccacc 677
    |||||||
Db 204 ATTGCATTAAGCGAAGCTGCTGTGAATTTGAGAGTGGCCGGAGAGATGGCCACCC 145
OY 678 cacatgaagaactctgtgaccaaagttcgggttgaggagttgtgggggaagtgaaagg 737
    |||||||
Db 144 CACATGAAGAACTTTGTGACCAAGCTTTGCGTGGGAGATTGTGGGGGAGCTGAAGGG 85
OY 738 aaaagaagaagaattccaagaanaaaatgcgcacatagctgcttcttgaagagcgaaga 795
    |||||||
Db 84 AAAAGCAAGAAATTTCAGAGAAATAATGCCGCAATAGCTGTCTGAGAGAGCTGAAGA 27

```

```

RESULT 11
LOCUS BE674157 616 bp mRNA EST 08-SEP-2000
DEFINITION 7d78a04.x1 NCI-CCRP_Lu24 Homo sapiens cDNA clone IMAGE:3278862 3'
    similar to TR:095793 095793 STRUFEN PROTEIN. ;, mRNA sequence.
ACCESSION BE674157
VERSION BE674157.1 GI:10034698
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
    1 (bases 1 to 616)
    NCI-CCRP http://www.ncbi.nlm.nih.gov/ncicrp.
    National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
    Tumor Gene Index
    Unpublished (1997)
    Contact: Robert Strausberg, Ph.D.
    Tel: (301) 496-1550
    Email: Robert.Strausberg@nih.gov
    Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
    Emmerit-Buck, M.D., Ph.D.
    cDNA Library Preparation: M. Bento Soares, Ph.D.
    cDNA Library Arrayed by: Greg Lennon, Ph.D.
    DNA Sequencing by: Washington University Genome Sequencing Center
    Clone distribution: NCI-CCRP clone distribution information can be
    found through the I.M.A.G.E. Consortium/ULNL, send email to:
    InfoImage.Lnl.gov
    Seq primer: -40UP from Gibco
    High quality sequence stop: 469.
    Location/Qualifiers
        1. 616
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:3278862"
        /clone_lib="NCI-CCRP_Lu24"
FEATURES
    source

```

```

/lisse_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: pF713D-Pac (Pharmacia) with a
modified polylinker. Plasmid DNA from the normalized
library NCI-CCRP_Lu24 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (clonoids
1414920-1417991 and 1520504-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 179 a 153 c 139 g 141 t 4 others
ORIGIN
Query Match 18.9%; Score 594.4; DB 111; Length 616;
Best Local Similarity 98.4%; Pred. No. 1.4e-135;
Matches 598; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
OY 1155 aagacacccataagaagaccagaggatggaagaagaataaccttttgaactgagct 1214
    |||||||
Db 8 AAGACACCCATTAAGAAACAGGAGATGCAAGAAAGTAACCTTTTGAACCTGCTCT 67
OY 1215 gggatgaaatlggaactagtaataagaagatgagttcaagatgacctataatgcat 1274
    |||||||
Db 68 GGGATGAATAATGGAGCTAGTATAAAGAGATGAGTTCAGATGCTTATCTAAGTCAAT 127
OY 1275 cagcagctgctgctggaattctcccatggtgcccgaagtgccgaagctgtagaagt 1334
    |||||||
Db 128 CAGCAGCTCCTGCTGGAATTTCTCCATGTTGTCGAGGTGCGCCAGCTGTAGAGATT 187
OY 1335 agtcaagagacatcacacaaagatttaccagaggagctccgaactctccaagggcaag 1394
    |||||||
Db 188 AGTCAAGGACATCACACCAAGATTTTACCAGGGCAGCTCCGATCTGCAAGGCCACG 247
OY 1395 gtaactgcatgataagcccgagagttgtgtatgaggggagacctgcccacagcgagacc 1454
    |||||||
Db 248 GTAACGTCCATGATAGCCGAGAGTTGTTATGGGGGACACCTGCCACAGCGGAGACC 307
OY 1455 attttaagaataaactctcttccagcagcagcagcagcagcagcagcagcagcagcagc 1514
    |||||||
Db 308 ATTTTAAAGAAATACATCTCTTCAGGCCAGCAATGACCTTCTCAAGACCTCTCT 367
OY 1515 gagcaactgactatcttccagagagtcagaggatcagagttcgaatacaagaagctcccg 1574
    |||||||
Db 368 GAGCACTGGACATCTTTCCAGAGTCCAGGATTTCCAGTTCAATACAAAGACTTCCCC 427
OY 1575 aaaaacaagaagaagaattglatctctatcaattgctcctcagccactgtgac 1634
    |||||||
Db 428 AAAAACAACAAGAAGCAATTTGATCTTTATCAATTTGCTCTCCTCANCACCTGTGATC 487
OY 1635 agccatggtatcggcagaagatgtagagctcctcagatgagtgctgctggaacatctta 1694
    |||||||
Db 488 AGCCATGATATGAGCAAGATGTGAGTCTCTGCAATATGAGTGGCTGGAACATCTTA 547
OY 1695 aagtgctgtcgtgagtgagcacaagaatgacagatgccaagaacaggaacagacca 1754
    |||||||
Db 548 GACTGTCTGTGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 607
OY 1755 atgtctgt 1762
    |||||||
Db 608 ATGTCTGT 615
RESULT 12
LOCUS AW162002 594 bp mRNA EST 09-NOV-1999
DEFINITION au72c03.x1 Schneider fetal brain 00004 Homo sapiens cDNA clone
    IMAGE:2781796 3' similar to TR:095793 095793 STRUFEN PROTEIN. ;,
    mRNA sequence.
ACCESSION AW162002
VERSION AW162002.1 GI:6301035
KEYWORDS EST.

```

SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 594)
AUTHORS	Hillier,L., Allen,M., Bowles,L., Dubnue,T., Gaisel,G., Jost,S., Kizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin J., Moore,B., Schellenberg,K., Stepien,M., Tan,F., Theisling,B., White,Y., Wyllie,T., Waterston,R. and Wilson,R. WashU-NCI human EST Project Unpublished (1997) Other ESTs: au72c03.y1 Contact: Wilton RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: estw@wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: -400P from Gibco High quality sequence stop: 464.
TITLE	
JOURNAL	
COMMENT	
FEATURES	
SOURCE	location/Qualifiers  1..594 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2781796" /clone_lib="Schneider fetal brain 00004" /sex="male" /tissue_type="frontal lobe" /dev_stage="5 months post-conception" /lab_host="DH10B" /note="Organ: brain; Vector: pBluescript SK (Stratagene); Site 1: SstI; Site 2: XhoI; Double-stranded cDNA was prepared from human fetal brain tissue. 5' and 3' adaptor adaptors were used in cloning as follows: 5' adaptor sequence: 5'-GAGAGAGAAGACCTCAGCATCTTAATTAATTAAATCCCCCCCCC-3' and 3' adaptor sequence: 5'-GAGAGAGACTCGATTTTGTTTTTTTTTTTT-3'. The library was size-selected for >0.5 kb inserts and has an average insert size estimated at 1.2 kb. This library was constructed using the CAP-trapper method for full-length enrichment and has not undergone amplification. Library was constructed by Dr. Claudio Schneider (UNCIB-Area Science Park, Trieste, Italy)." BASE COUNT       130 a     152 c     123 g     188 t           1 others ORIGIN
Query Match	18.7%; Score 588.2; DB 40; Length 594;
Best Local Similarity	99.3%; Pred. No. 4.7e-134;
Matches 590; Conservative	0; Mismatches 4; Indels 0; Gaps 0;
OY	249 gaaagcataaccctactgtagaacataatbcacgylcatgaanaactggaaaaaacca 308 
Db	594 GAACACATAAACCCCTACGCTGAAGACTAATATGCACGTCGTCAATCTTGAAAAAACCCA 535 
OY	309 atgtataagccgtgttgcaccttactcttcgatgtaagtcacccaactaactacaatgsa 368 
Db	534 ATGTATTAAGCCGTGTGACCCCTTACTCTCGAGATGCAGTCCACCCTTAATACATGATGAGA 475 
OY	369 ggagtgctttcccccaaggtaactttaccatttccagtccacttacttatcaa 428 
Db	474 GGAGTGCTTATCCGCCCAAGGACTTTTAAACCATTTCCAGTTCCACCTTTACTTATCAA 415 
OY	429 gtggaacttctgtgtggagaagacaatatltatggcaagaagaaaagacaagacgtcg 488 
Db	414 GTGGAACTTTCGTGTGGAGNAGACGAATTTAATATGCAAAAAGAAACAAGACAGCTCG 355 
OY	489 aaacacgaltgctgtgccaaagcgttggagatcctgcgaatlgagccctctgcaagag 548 
Db	354 AAACACGATGTGCTGCCCAAGCGCTTGGAGATCCCTGCAGAAATGAGACCCCTGCCAAGAG 295 

QY	549	ctgagtggaattggaagaagaaatccgaagaagaagaatccaaataatctgaaataagca	608
Db	234	CTGAGAGTGGAATGCAAGAGAAATCCGAGACAAATAATCTCAATTAATCTGAATTAAGTCA	235
QY	609	gtgtttggaatttgacttaaacggaacttgccctgtgaaatttcgaagtgtgcccggagaagt	668
Db	224	GTGTTTGGAAATTTGCACTTAACCGGAACCTTGCCTGTGAATTTTCGAGTGCGCCCGGAGAGT	175
QY	669	ggcccaacccccacatgaagaactttgttcacaaagtttcggttggggagtttgttggggaa	728
Db	174	GGCCACACCCCAATGAAGAACTTTGTGTGACCAAGTTTCGTTGGGGAGTTGTGTGGGGAA	115
QY	729	gttgaaggagaaagaagaagaatttcaagaagaagaatgcgcgcataagctgttcttgagaag	788
Db	114	GGTGAAGGGAAGCAAGCAAGAAATTTCAAGAGAAATATGCCGCATAGCTGTCTTGAAGAG	55
QY	789	ctgaagaagttacacgcccctgcctgcagtttgaacgagtaaaccttagaatcaaa	842
Db	54	CTGAAGAAGTTACCGCCCTCTGCTGCACATTTACAGAGATAAAGCCTAGAAAAAA	1
RESULT	13		
LOCUS	AV695897	612 bp	mRNA EST 25-SEP-2000
DEFINITION	AV695897 GKC Homo sapiens cDNA clone GRCER08 5', mRNA sequence.		
ACCESSION	AV695897		
VERSION	AV695897.1	GI:10297760	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi; Mammalia: Eutheria: Primates: Catarrhini; Homnidae: Homo. 1 (bases 1 to 612)		
AUTHORS	Wu,T., Qian,B., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H., Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.		
TITLE	Homo sapiens cDNA GK- clones		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Zeguang Han Chinese National Human Genome Center at Shanghai 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203. P. R. China Tel: 86-21-50801919(ex.45) Fax: 86-21-50801922 Email: hanzg@chgc.sh.cn This clone is available at CHGC in Shanghai.		
FEATURES	Location/Qualifiers		
SOURCE	1. 612		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="GRCER08"		
	/clone_1id="GKC"		
	/tissue_type="hepatocellular carcinoma"		
	/dex_stage="Adult"		
	/lab_host="SOLR"		
	/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"		
BASE COUNT	149 a 164 c 125 g 174 t		
ORIGIN			
Query Match	18.2%; Score 571.4; DB 38; Length 612;		
Best Local Similarity	98.4%; Pred. No. 6.5e-130;		
Matches 611; Conservativity	0; Mismatches 1; Indels 9; Gaps 3;		
QY	1988	gttagacatcggtccctccacgaatagacgtgcctgtggtgggaagaaggtccccaacc	2047
Db	1	GGTGACCTGGGCGGTCCCTACGCAATAGCAGGCTGCTGTGGGGAACAAGGCTGCCACGC	60
QY	2048	agctggttctcccggaacacagacatcacacacctgggaactcgttgtttggtctt	2107





KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 609)  
AUTHORS Dias Neto, E., Garcia Correia, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Baia, G. S., Simpson, D. H., Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J. J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-WR4-PT0052-220500-206-f06&t3=2000-05-22&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 10  
High quality sequence stop: 589.  
Location/Qualifiers  
1. 609  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="PT0052"  
/dev\_stage="Adult"  
/note="Organ: pnet; Vector: puc18; Site\_1: Sma1; Site\_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
BASE COUNT 173 a 153 c 145 g 138 t  
ORIGIN

Query Match 17.7%; Score 555; DB 136; Length 609;  
Best Local Similarity 97.9%; Pred. No. 7.1e-126;  
Matches 594; Conservative 0; Mismatches 10; Indels 3; Gaps 3;

QY 1222 aaaaaggactagtaataaaggatgagtgatgccttatcaatgcatcagcagc 1281  
DB 6 AACATGGGACTAGTAAT-AAGAGAGATGAG-TCAGGATGCTTATCTAAGTCATCAGCAGC 63  
QY 1282 tggcctcgtgaatcttcacatagtgcccaaggctgcgcacagctgtagaagttagaag 1341  
DB 64 TGCCTGCT-GAATTCCTCCATGCTGCGCCGAGGTGCGCGGCTGTAGAGTTAGTCAAG 122  
QY 1342 gacatacaccaagaatttaccaaggcagctcgaatcctgcacaaggccacgtaactg 1401  
DB 123 GACATCACACCAAGATTTCACAGGGCAGCTCGAATCCTGCGCAGGGCCACGGTAAC TG 182  
QY 1402 ccattatagccgagagtggtgtatgggggcaacctgcgccacaagccagacatttaa 1461  
DB 183 CCATGATAGCCCGAGAGTAGTGTATGGGGCAGCTCGCCACAGCCGAGACCATTTTAA 242  
QY 1462 agataaacatctcttcagagcagctacccatgacctctcagagagaccccttgagcaac 1521  
DB 243 AGAATAACATCTCTTCAGGGCAGCTACCCATGGAGCTCTCAGCAGACCTCTGAGCAAC 302  
QY 1522 tggactactcttcacagagtlccaggagtlcgaagttgaatacaaaagacttcccaaaaaa 1581

DB 303 TGGACTATCTTTCCAGAGTCCAGGGATTCGAGGTGAATTAACAAGACTTCCCAAAAACA 362  
QY 1582 acaagaagaatttgratcttataatgctccctcagccacctctgataagccatg 1641  
DB 363 ACAAGAACGAAATTGTATCTTATCAATGCTCTCAGCCACCTCTGATCAGCCATG 422  
QY 1642 gtaacgcaagatctggaagtcctgcacatgatatgctgcgctgaacatcttaagttgc 1701  
DB 423 GTATCGGCAAGATGTGAGACTCTGCCATGATATGGCTGGCTGAGAACATCTTAAAGTTG 482  
QY 1702 tgcctgaagttgacccaagaagatcacagagatgcacaaagaagaacggaacatgtctg 1761  
DB 483 TGTCTAGTGTAGACCAACAAGATGACAGATGCCAAGAACGAGAACGCAATGTCTG 542  
QY 1762 tgggtggagtgctgaaccttctcggcactgaaccattataatcccaatatata 1821  
DB 543 TGTGTGGAGGTGCTGAACCTTTCTGGGATGTAACCATTTAATAATCTCAACATATATA 602  
QY 1822 ctgaaaa 1828  
DB 603 CTGGCAA 609

Search completed: April 5, 2001, 00:29:19  
Job time: 53055 sec

Fri Apr 6 15:25:24 2001

us-09-316-048-1.rst

---

Page 14



409 ATGCTCAAGTTCAGAGTTCAGAACCCATCTGCTCTCAGG 458  
 17 YserGlnIleuAsnLysAsnGlnSerIleuSerGlnProIleuMetS 34  
 459 GAGCCAAATGACGACAGACAGAGCTCTCTCTCAGCAGCCTTATGA 508  
 34 exIleProSerThrThrSerSerIleuProSerGluAsnIleuGlyArgPro 50  
 509 GATATCCTTCTACTACTACTACTCTCTGCTGAAATGCAAGTATGACCC 558  
 51 IlegIlnAsnSerAlaIleuProSerAlaSerIleThrSerThrAlaAl 67  
 559 ATTCAAAATCTGCTTTACCTCTCTGATCATTTACATCCACCGAGTCAGC 608  
 67 aIaIleuSerIleThrProThrValGluIleuAsnAlaIleuCysMetLysI 84  
 609 TGCAGAAACATTAACCCCTACTCTGTAGACATAAATGCACTGTGATTAAC 658  
 84 euGIYlYsLysProMetYlYsLysProValAspProTyrSerArgMetGln 100  
 659 TTGGAATAAAACCAATGATTAAGCCTGTGACCTTACTCTGAGATGACAG 708  
 101 SerThrTyAsnTyAsnMetArgGlyGlyAlaTyrProProArgTyrPh 117  
 709 TCCACCTTAACCTACACATGAGAGAGAGGTGCTTATCCCCGAGTACTT 758  
 117 eTyTyrProPheProValProProIleuTyrGlnValGluIleuSerValG 134  
 759 TTACCATTTCCAGTTCACCTTTACTTTATCAAGTGAACCTTCTGTGG 808  
 134 YglYglngInPheAsnGlyLysGlyLysThrArgGlnAlaIleuHis 150  
 809 GAGGACAGCAATTTATGGCAAAAGAAACAAAGACAGCGCTCGCAAAAC 858  
 151 AspAlaIleAlaLysAlaIleuArgIleuGlnAsnGlnProIleuProG 167  
 859 GATGCTGCTGCCAAAGCTTTGAGATCTGCAAGATGACCCCTGCCAGA 908  
 167 uArgIleuGlnValAsnGlyArgGluSerGluGluAsnLysLys 184  
 909 GAGCGTGGAGGTGAATGGAAGAAATCCGAAAGAAATCTCAATAAAT 958  
 184 eGluIleSerGlnValPheGluIleAlaIleuLysArgAsnLeuProVal 200  
 959 CTGAATTAAGTCAAGTGTGAGATGCACTTAACGGAACCTGCTGTG 1008  
 201 AsnPheGluValAlaIleuArgIleuSerGlyProProHisMetLysAsnPh 217  
 1009 AATTTCGAGGTGGCCCGGAGAGGTGGCCACCCACATGAAAGAACTTTGT 1058  
 217 IThrLysValSerValGlyIlePheValGlyGluGlyLysSerL 234  
 1059 GACCAAGGTTTCGTTGGGAGTTTGTGGGGAAGTGAAGGAAAGCA 1108  
 234 YslYslIleSerLysLysAsnAlaAlaIleAlaValLeuGluIleuLys 250  
 1109 AGAAGATTTCAAAAGAAATGCGCGCATAGCTGTCTGTGAGAGCTGAAG 1158  
 251 LysIleuProIleuProAlaValAlaGluArgValLysProArgIleLys 267  
 1159 AAGTTACCCCTGCTGCTGAGTGAAGAGTAAACCTGAAATCAAAA 1208  
 267 sTyThrLysProIleValLysProGlnThrSerProGluTyrGlyGlnG 284  
 1209 GAAACAAAACCATGATGACAGCCACAGCAAGCCCAATATATGGCCAGG 1258  
 284 YtIleAsnProIleSerArgLeuAlaGlnIleGlnIlnAlaLysGlu 300  
 1259 GGATCAATCCGATTAAGCCAGTGGCCAGATCCAGCAGCAAAAGAGAG 1308  
 301 LysGluProGluTyrThrIleuThrArgIleuArgGlyLeuProArgArg 317  
 1309 AAGGAGCCAGATACAGCTCTCTCAAGAGAGGAGGCTCCGCGCGCAG 1358

317 gGluPheValMetGlnValLysValGlyAsnHisThrAlaGluGlyThrG 334  
 1359 GAGTTTGTGATGACAGGTGAGGTTGAAACCACTGCAAGAAAGAACGG 1408  
 334 YThrAsnLysLysValAlaLysArgAsnAlaIleuAsnMetLeuGlu 350  
 1409 GCACCAAAAGAGGTGGCCAAAGCCCAATGACAGCGAGAAATGCTGGAG 1458  
 351 IleuGlyPheLysValProGlnArgGlnProThrLysProAlaLeu 367  
 1459 ATCTTGGTTTCAAAATCCCGCAGCGGAGCCACCAAAACCCGACTCA 1508  
 367 sSerGluGluLysThrProIleLysLysProGlyAspGlyArgValT 384  
 1509 GTCAGAGGAGAGACACCCATTAAGAAACAGGAGGTGGAAGAAAGTAA 1558  
 384 hrPhePheGluProGlySerGlyAspGluAsnGlyThrSerAsnLysGlu 400  
 1559 CTTTTTTGAACCTGCTGCTGGGATGAAATGGAGTAAATGATAAAG 1608  
 401 AspGluPheArgMetProTyrIleuSerHisGlnGlnIleuProAlaGly 417  
 1609 GATGAGTTAGAGATGCTTATTAAGTCATCAGCACCTGCTGTGAAT 1658  
 417 eLeuProMetValProGluValAlaGlnAlaValGlyValSerGlnGly 434  
 1659 TCTTCCCATGTCGCCGAGGTGCCAGGCTGAGAGATGATGACAGAC 1708  
 434 IShIsthrLysAspPheThrArgAlaAlaProAsnProAlaLysAlaThr 450  
 1709 ATCCACACCAAGATTTTACCAGGCGAGCTCCGAAATCTCCAAAGGCAAG 1758  
 451 ValThrAlaMetIleAlaArgGluIleuLeuTyrGlyLysThrSerPro 467  
 1759 GTRACTGCCATGATAGCCGAGAGTTGTGTATGGGGACCTCGCCAC 1808  
 467 rAlaGluThrIleLeuLysAsnAsnIleSerSerGlyHisValProHisG 484  
 1809 AGCGAGACCATTTTAAGATTAACATCTCTTCAGGCCAGCATGCCCATG 1858  
 484 LyrProLeuThrArgProSerGluGlnLeuAspTyrLeuSerArgValGln 500  
 1859 GACCTTCACGAGACCTCTGAGCAACTGACATATCTTCCAGAGTCCAG 1908  
 501 GlyPheGluValGluTyrLysAspPheProLysAsnAsnLysAsnGluPh 517  
 1909 GGATTCAGAGTGAATTAACAAGACTTCCCAAAACAAACAAAGAGCAAT 1958  
 517 eValSerLeuIleAsnGlySerSerSerGlnProProLeuIleSerHisGly 534  
 1959 TGTATCTCTTAACAATTTGCTCTCTCAGCACCTCTGATCAGAGCAATG 2008  
 534 IeGIYlYsAspValGluSerCysHisAspMetAlaIleuAsnIleu 550  
 2009 TCGGCAAGGATGTGGAGTCTGCTGATGATATGCTGCTGAAACATCTTA 2058  
 551 LysIleuSerGluIleuAspGlnIleuSerThrGluMetProArgThrG 567  
 2059 AAGTTCTGTCTGAGTTGACCAACAAGTACAGAGATGCCAAGAACAGG 2108  
 567 YAsnGlyProMetSerValCysGlyArgCys 577  
 2109 AAACGAGCAATGTCTGTGTGGAGGTGC 2139

seq\_name: gb\_pf3:AF061940  
 seq\_documentation\_block: 3142 bp mRNA PRI 04-MAR-1999  
 LOCUS AF061940  
 DEFINITION Homo sapiens stauflen protein (STAU) mRNA, alternatively spliced,  
 complete cds.  
 ACCESSION AF061940  
 VERSION AF061940.1 GI:4335950



```

398 snlysgluaspgluphearqmetprotyrleuSerHisInglInleuPro 414
|||||
1237 ATAAAGAGATGAGTTCAGAGATGCCCTATCTATCAATCATACAGACGCGCT 1286
|||||
415 AlaglylleuPrometValProgluValAlaglnAlaValgylValSe 431
|||||
1287 GCTGGAATTCCTCCATGCTGCCGAGGCTGCCAGGCTGTAGAGACTTAG 1336
|||||
431 rclInglYHisHsthThrLysAspPheThrArgAlaAlaProasnProAlaL 448
|||||
1337 TCAAGGACATCAACACCAAGATTTCACAGGCGAGCTCCGAATCCGCA 1386
|||||
448 yValglnValThrAlaMetIleAlaArgIleuLeuTyrcylgylThr 464
|||||
1387 AGGCCACGTAAGTCCATGATGATGCCCGAGAGTGTGTATGGGGCACC 1436
|||||
465 SerProThrAlaGluThrIleLeuLysAsnAsnIleSerSerglyHisVa 481
|||||
1437 TCGCCACACGCGAGCACATTTTAAAGATAACATCTCTTCAGGCGACGT 1486
|||||
481 lProHsglyProleuThrArgProSergInglInleuAspTyrcylSerA 498
|||||
1487 ACCCATGAGACCTCTCAGGACCCCTGAGCAACTGAGACTATCTTCCA 1536
|||||
498 rValglnGlnlyPheGlnValGluTyrcylLysAspPheProLysAsnLys 514
|||||
1537 GAGTCCAGGAGTTCAGAGTTCGAATACAAAGACTTCCCCAAAACAACAG 1586
|||||
515 AsnGluPheValSerleuIleasnYssSerSergInProleuIleSe 531
|||||
1587 AACGAAATTTGATCTCTATCAATGCTCTCTCAGCCACTCTGTATCAG 1636
|||||
531 rHisGlyIleGlyLysAspValGluSerCysHisaspMetIleAlaLeuA 548
|||||
1637 CCATGGTATCGCAAGAGATGGAGTCCGCAATGATATGGCTGGCTGA 1686
|||||
548 snIleLeuLysLeuLeuSerGluLeuAspGlnGlnSerThrGluMetPro 564
|||||
1687 ACATCTTAAAGTTGCTGCTGTGAGTTGGACCAACAAGTACAGAGATGCA 1736
|||||
565 ArgThrGlyAsnGlyPrometSerValCysGlyArgCys 577
|||||
1737 AGAACAGGAAACGACCAATGCTGTGTGGAGGTGC 1775
|||||
seq_name: gb_Dr3:AF061938

seq_documentation_block:
LOCUS AF061938 3217 bp mRNA PRI 04-MAR-1999
DEFINITION Homo sapiens staufen protein (STAU) mRNA, alternatively spliced,
complete cds.
ACCESSION AF061938
VERSION AF061938.1 GI:4335946
KEYWORDS
SOURCE
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 3217)
AUTHORS Wickham,L., Duchaine,T., Luo,M., Nabl,I.R. and Desgroselliers,L.
Mammalian staufen is a double-stranded RNA and tubulin binding
protein which localizes to the rough endoplasmic reticulum
Mol. Cell. Biol. 19 (3), 2220-2230 (1999)
JOURNAL
MEDLINE 99147057
REFERENCE 2 (bases 1 to 3217)
AUTHORS Desgroselliers,L., Wickham,L. and Luo,M.
TITLE Direct Submission
JOURNAL Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900
Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada
FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

```

/chromosome="20"
/map="20q13.1"
1..3217
/gene="STAU"
363..1853
/gene="STAU"
/function="RNA binding protein"
/note="alternatively spliced transcript"
/codon_start=1
/product="staufen protein"
/protein_id="AADI7530.1"
/db_xref="GI:4335947"
/translation="MKLGKKPKMKPVDPYSRMOSTYNNMKGAVPPRYFPPEPPL
LYVELSVGGQOFNGKGTROAAKHDAAKARILQNEPLPERLEVNGRESSEENLNK
SEISOVEILAKRLNLPVNFYARSGPBNKFNFTKYSVGEVGEKSKISKNA
AIALVELKRLPLPAVERYKPRTKTKTKPLVKRQSPPEYGGGNIPTSLAIOQAKK
EKPEFTLTREKLPKRREYVQYKVNHTAEIGTKTKKAKRAAEEMLEIGKRP
QROPTKALKSEKTPIKRPDGRKVTFFEGSDGENTSKKEDFRMPYLHQDPA
GILPMVEVAQAVSOGHHTKDFTRAAPNPAKATYAMTARELLYGTSPTAETILK
NMTSGHVPKPLTRPSEOLDYLSRVQGFQVEYKDFPKNNKNEFVSLINCSOPPLIS
HGICKDVESCHDMAALNLIKLSLELDQOSTEMPRTGNGPMSVCGRC"
BASE COUNT 889 a 796 c 742 g 790 t
ORIGIN

alignment_scores:
Quality: 2659.00 Length: 577
Ratio: 4.906 Gaps: 3
Percent Similarity: 93.934 Percent Identity: 91.508

alignment_block:
US-09-316-048-6 x AF061938 ..
Align seg 1/1 to: AF061938 from: 1 to: 3217
15 LeuSerGlySerGlnIleLeuAsnLysAsnGlnSerleuLeuSerGlnPr 31
||| ||| ::| ::| ::| ::| ::| ::| ::|
121 CTCCAGCGGCGACTCCGCCCTTCCTCCCTGCTCCCTCTCTCTCTCC 170
||| ::| ::| ::| ::| ::| ::| ::|
31 oleuMetSerIleProSerThrThrSerSerleuProSerGluAsnAlag 48
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
171 TTTTTCCTTCTCTCTCTCCCTCCCTCTCCGCGCCGCGCCAGACGGCG 219
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
48 lYArg.....ProleglnAsn.....SerAlaLeuPro 57
||| ||| ||| ::| ::| ::| ::|
220 GCCGGGAGCAGAGTCCGGAGCAGACCCAGAGTTTATTACCACTTAACC 269
||| ::| ::| ::| ::| ::| ::| ::|
58 SerAlaSerIleThrSerThrSerAla.....Al 67
||| ::| ::| ::| ::| ::| ::| ::|
270 TCTCAGAACTGAACAAAGACAAACATGTTCTCTGAAGCCCTCTTTTAA 319
||| ::| ::| ::| ::| ::| ::| ::|
67 alaGluSerIleThrProThrValGluLeuAsnAlaLeuCysMetLysL 84
|||||
320 AAAAGAAACATTAACCCCTACTGTAGAACTAAATGACATGTCATGAAC 369
|||||
84 euGlyLysSerPrometTyrcylLysProValAspProtyrSerArgMetGln 100
|||||
370 TTGGAAAAAACCAATGTAAGCTGTGACCTTACTCTGGAGTACGAG 419
|||||
101 SerThrTyrcylLysAsnMetArgGlyAlaTyrcylProProArgTyrcpH 117
|||||
420 TCCACTCTAATACACACTGAGAGAGGTGCTTATCCCGCAGGACTT 469
|||||
117 eTyrcProPheProValProProleuLeuTyrcylValGluLeuSerValG 134
|||||
470 TTACCAATTTCCAGTTCCACTTCTTATCAAGTGAACCTTCTGTGG 519
|||||
134 TyrcylGlnGlnPheAsnGlyLysGlyLysThrArgGlnAlaAlaLysHis 150
|||||
520 GAGGACAGCAATTTAATGCAAGAAAGACAAAGACAGGCTCGAAACAC 569
|||||
151 AspAlaAlaAlaLysAlaLeuArgIleLeuGlnAsnGluProleuProcl 167
|||||

```

570 GATGCTGCTGCCAAAGCGTTGAGGATCTGCAGAAATGAGCCCTGCCAGA 619  
 167 uATgLeuGlValAsnGlyArgGluSerGluGluAsnLeuAsnLys 184  
 620 GAGGCTGAGGTGAATGAGAGAGATCCGAGAGAAATCTCAATTAAT 669  
 184 eGluLeuSerGluValPheGluIleAlaLeuLysArgAsnLeuProVal 200  
 670 CTGAATAATGATCAAGTGTGAGATTGCACTTAACGAACTTGCCGTG 719  
 201 AsnPheGluValAlaArgGluSerGlyProGluHisMetLysAsnPheVal 217  
 720 AATTGAGAGTGGCCCGGAGAGTGGCCCAACCATGAAAGAACTTTGT 769  
 217 LThrLysValSerValGlyGluPheValGlyGluGlyGlySerL 234  
 770 GACCAAGGTTTCGTTGGGAGTTGTGGGGAAGTGAAGGGAAGCA 819  
 234 ySLysLLeSerLysLysAsnAlaAlaIleAlaValLeuGluLys 250  
 820 AGAAGATTTCAAGAAATGCGCCCATACCTGTTCTTGAGAGCTGAC 869  
 251 LysLeuProLeuProAlaValAlaArgValLysProArgIleLysLy 267  
 870 AAGTTACCGCCCTGCTGCTGAGTTGAACGAGTAAAGCCTAGAAATCAAAA 919  
 267 sLysThrLysProLLeValLysProGluThrSerProGluTyrGlyGln 284  
 920 GAAAGCAAAACCATAGTCAAGCCACAGACACCCAGAAATATGGCCAG 969  
 284 LylLeAsnProLLeSerArgLeuAlaGlnIleGlnAlaLysLysGlu 300  
 970 GGATCAATCGATTAGCCGACTGGCCAGATCCAGCGAGAAAGAGAG 1019  
 301 LysGluProGluTyrThrLeuLeuThrGluArgGlyLeuProArgArg 317  
 1020 AAGGAGCAGAGTACACGCTCTCACAGAGCGAGGCTCCCGCCGCGAG 1069  
 317 gGluPheValMetGlnValLysValGlyAsnHisThrAlaGluGlyArg 334  
 1070 GGGATTGTGATGACAGGTGAAGTTGAAACCACTGCAGAAAGAAACGG 1119  
 334 LylThrAsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGlu 350  
 1120 GCACCAACAGAGAGTGGCCAGCGCATGCGAGCGAAGAACATGCTGAG 1169  
 351 IleLeuGlyPheLysValProGlnArgGlnProThrLysProAlaLeu 367  
 1170 ATCCTTGTTCAAAAGTCCCGCAGCGCACCCCAACCCGCACTCAA 1219  
 367 sSerGluGluLysThrProLLeLysLysProGlyAspGlyArgLysValT 384  
 1220 GTGAGGAGAAAGACCCATTAAGAAACAGGAGATGAGAAAGATAA 1269  
 384 hPhePheGluProGlySerGlyAspGluAsnGlyThrSerAsnLysGlu 400  
 1270 CTTTTTTGAACCTGGCTCGGGATCAAAATGGGACTAGTAATAAAGAG 1319  
 401 AspGluPheArgMetProTyrLeuSerHisGlnGlnLeuProAlaGly 417  
 1320 GATGAGTTTCAGGATGCTTATCTAAGCATCAGACGCTGCTGGAAT 1369  
 417 eLeuProMetValProGluValAlaGlnAlaValGlyValSerGlnLys 434  
 1370 TCTTCCCATGTCGCGAGGTCCGCGAGGCTGTAGGAGTTAGTAAGGAC 1419  
 434 IsthIsthLysAspPheThrArgAlaAlaProAsnProAlaLysAlaThr 450  
 1420 ATCAGACAAAGATTTTACAGGCGACCTCGAATCTGCCAAGCCACG 1469  
 451 ValThrAlaMetIleAlaArgGluLeuLeuTyrGlyLysThrSerProTh 467  
 1470 GTACTGCTCATGTAGCCGAGAGTTGTGTATGGGGGACCTGCCAC 1519

467 rAlaGluThrIleLeuLysAsnAsnLLeSerSerGlyHisValProHisG 484  
 1520 AGCGAGACCATTTTAAAGATTAACATCTCTTCAGGCGACGATCCCATG 1569  
 484 LProLeuThrArgProSerGluGlnLeuAspTyrLeuSerArgValGln 500  
 1570 GACCTTCAGAGACCTCTGAGCACTGAGCATATCTTCCAGAGTCCAG 1619  
 501 GlyPheGlnValGluTyrLysAspPheProLysAsnLysAsnGluPh 517  
 1620 GGATTCAGGTTGATACAAAGACTTCCCAAAAACAACAGAAACGAAAT 1669  
 517 eValSerLLeAsnCySerSerSerGlnProProLeuIleSerHisGly 534  
 1670 TGTATCTTATCATTTGCTCTCTCTCAGCCACTTCTATCAGCATGTA 1719  
 534 LeGlyLysAspValGluSerCyHisAspMetAlaAlaLeuAsnIleLeu 550  
 1720 TCGGCAAGGATGTGAGTCCGCGCATGATATGGCTGCGTGAACATCTTA 1769  
 551 LysLeuLeuSerGluLeuAspGlnGlnSerThrGluMetProArgThr 567  
 1770 AAGTTGCTGTCTGAGTTGACCAACAAAGTACAGAGATGCCAAGACAG 1819  
 567 yAsnGlyProMetSerValCySerGlyArgCyS 577  
 1820 AAGCGACCAATGCTGTGTGTGGAGGTGC 1850

seq\_name: gb\_p13:AF061941

seq\_documentation\_block: 3349 bp mRNA PRI 04-MAR-1999  
 LOCUS AF061941  
 DEFINITION Homo sapiens staufen protein (STAU) mRNA, alternatively spliced,  
 complete cds.  
 ACCESSION AF061941  
 VERSION AF061941.1 GI:4335952  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 WICKHAM.L., DUCHAINE.T., LUO.M., NABI.I.R. and DESGROSELLERS.L.  
 Mammalian staufen is a double-stranded RNA and tubulin binding  
 protein which localizes to the rough endoplasmic reticulum  
 Mol. Cell. Biol. 19 (3), 2220-2230 (1999)  
 JOURNAL 99147057  
 MEDLINE 2 (bases 1 to 3349)  
 REFERENCE Desgroseillers.L., Wickham.L. and Luo.M.  
 Direct Submission  
 Submitted (29-APR-1998) Biochemistry, University of Montreal, 2900  
 Edouard Montpetit, Montreal, Quebec H3T 1J4, Canada  
 JOURNAL Location/Qualifiers  
 FEATURES  
 source 1..3349  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="20"  
 /map="20q13.1"  
 1..3349  
 /gene="STAU"  
 /gene="STAU"  
 /gene="STAU"  
 /function="RNA binding protein"  
 /note="alternatively spliced transcript"  
 /codon\_start=1  
 /product="staufen protein"  
 /protein\_id="ABD1753.1"  
 /db\_xref="GI:4335953"  
 /translation="MKLGKKPMYKPYDPYSRMQSYTYNNMKGATPPRYFPYPPPL  
 LTVDELVSQGGQFNGKTRQAKHDAAKRALRIIONEFLPERLVNGRESSEENLNK  
 SEISQVFETALRLNLPVNFVARRESGPPHMKNFVYKVSVEGVEGEGSKITSKNA  
 AIAVLEIKLPLPAVERVKPRIKKTKPIVKPOTSPYGGGINDPISRLAIOQAKK

EKEPEYTLTERGLPRRREFVMOYKVGNHTEGTGNTKKVAKRNAENMLEILGPKVP  
QROPTKALKSEKTPIKKPGDKRVTFFEPGSDENGTSKDEFPMPYLSHQLP  
GIIPMEVAOAVCSOQHHTKDETPRAAPNPAKTVAMLRRELLYGTSTPATILK  
NNISSGHVPHGLRSPBOLDYLSRVQFOVEYEDFPKNNKNEFVSILINCSOPPLIS  
HGICKDYESCNDMALNLIKLSLSDOSTEMPTGMPKNSVCGRG

BASE COUNT 913 a 838 c 777 g 821 t  
ORIGIN

## alignment\_scores:

Quality: 2649.00 Length: 509  
Ratio: 3.204 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-316-048-6 x AF061941 ..

Align seg 1/1 to: AF061941 from: 1 to: 3349

69 GluSer1LeThrProThrValGluLeuAsnAlaLeuCysMetLysLeuG1 85  
|||||  
456 GAAAGCATACCCCTACTGTAGACATAAATGCATCTGCATGAACCTTG 505  
85 yLysLysProMetTyrLysProValAspProTyrSerArgMetGlnSerT 102  
|||||  
506 AAAAAACCAATGTATAGCTGTGACCTTACTCTCGATGACATGCCA 555  
102 hTyrTrsntYrAsnMetArgGlyAlaTyrProProArgTyrPheTyr 118  
|||||  
556 CCTATACTACCAACATGAGAGGAGGTATATCCCGAGGATACCTTTTAC 605  
119 ProPheProValProProLeuLeuTyrGlnValGluLeuSerValG1 135  
|||||  
606 CCATTTCCAGTTCACCTTACTTATCAAGTGAAGCTTTCTGTGGAGG 655  
135 yGlnGlnPheAsnGlyLysGlyLysThrArgGlnAlaAlaLysHisAspA 152  
|||||  
656 ACAGCAATTTAATGGCAAGAAAGACAAGACAGCTCGCAAAACACGATG 705  
152 lAlaAlaLysAlaLeuArgGlyLeuGlnAsnGluProLeuProGluArg 168  
|||||  
706 CTGCTGCCAAAGCGTTGAGATCTCTGCAAAATGAGCCCTGCCAGAGAG 755  
169 LeuGluValAsnGlyArgGluSerGluGluGluAsnLeuAsnLysSerG1 185  
|||||  
756 CTGGAGGTAAATGGAAGACAAATCCGAAGAAATCTCAATTAATCTGA 805  
185 uLysSerGlnValPheGluLeuAlaLeuLysArgAsnLeuProValAsnP 202  
|||||  
806 AATTAAGTCAAGTGTGAGATTGACACTTAACGGAACCTTGCTGTGAAT 855  
202 heGluValAlaArgGluSerGlyProProHisMetLysAsnPheValThr 218  
|||||  
856 TCGAGGTGGCCGCGAGAGTGGCCACCCACATGAAGAACTTTGTGACC 905  
219 LysValSerValGlyLupheValGlyGluGlyGluGlyLysSerLysLys 235  
|||||  
906 AAGGTTTCGTTGGGAGTTTGTGGGGGAAGGTGAAGGGCAAAAGCAAGA 955  
235 sLysSerLysLysAsnAlaAlaAlaValLeuGluGluLysLysLys 252  
|||||  
956 GATTTCAAGAAATAATGCCCATACCTGCTCTTGAGGAGCTGAAGAAGT 1005  
252 euProProLeuProAlaValGluArgValLysProArgGlyLeuLysLys 268  
|||||  
1006 TACCGCCCTGCTGCTGCAATTGAACGAGTAAAGCCATGAATCAAAAAA 1055  
269 ThrLysProLLeValLysProGlnThrSerProGluTyrGlyGlnGlyL 285  
|||||  
1056 ACAAAAACCATAGTCAAGCCACAGACCAAGCCAGAAATATGGCCAGGGGAT 1105  
285 eAsnProLLeSerArgLeuAlaGlnLeuGlnAlaLysLysGluLysG 302  
|||||

1106 CAATCCGATTAGCCGACCTGGCCAGATCCAGCAGGCAAAAAAGAGAGG 1155  
302 lUpProGluTyrThrLeuLeuThrGluArgGlyLeuProArgArgValGlu 318  
|||||  
1156 AGCCAGAGTACACGCTCTCTCACAGAGCAGAGGCTCCCGCGCCAGGGAG 1205  
319 PheValMetGlnValLysValGlyAsnHisThrAlaGluGlyThrGlyTh 335  
|||||  
1206 TTTGTGATCAGAGTGAAGGTTGGAAACCACTGCAGAAAGAAAGGCGCAC 1255  
335 rAsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGluLeu 352  
|||||  
1256 CAACAGAAAGGTGGCCAGCCAGCATGACCCGAGAACATGCTGGAGATCC 1305  
352 euGlyPheLysValProGlnArgGluProThrLysProAlaLeuLysSer 368  
|||||  
1306 TTGGTTTCAAAAGTCCCGCAGCGGCGACGCCACCAACCCGACCTCAAGTCA 1355  
369 GluGluLysThrProLLeLysLysProGlyAspGlyArgLysValThrPh 385  
|||||  
1356 GAGGAGAGACACACCCATTAAGAAACAGGAGATGGAAGAAAGTAACTTT 1405  
385 ePheGluProGlySerGlyAspGluAsnGlyThrSerAsnLysGluAspG 402  
|||||  
1406 TTTTGAACCTGGCTGCGGAGATGAATAATGGACTACTAATAAGAGGATG 1455  
402 lUpheArgMetProTyrLeuSerHisGlnGluLeuProAlaGlyLeuLeu 418  
|||||  
1456 AGTTCAAGATGCCCTTATCTTAAGTCATACAGCTGCTGCTGGAATTCCT 1505  
419 ProMetValProGluValAlaGlnAlaValGlyValSerGlnLysHisH 435  
|||||  
1506 CCCATGTCGCCGAGGTGCGCCAGGCTGAGAGTATAGTCAAGAGCATCA 1555  
435 sThrLysAspPheThrArgAlaAlaProAsnProAlaLysAlaThrValT 452  
|||||  
1556 CACCAAGATTTTACCAGGCGAGCTCCGAATCTCGCAAGGCGCACCGTAA 1605  
452 hValMetLeuAlaArgGluLeuLeuTyrGlyLysSerProThrAla 468  
|||||  
1606 CTGCCATGATAGCCGAGAGATTGTATAGGGGCACTGCGCCACAGGCC 1655  
469 GluThrLLeuLeuLysAsnAsnLLeSerSerGlyHisValProHisGlyPr 485  
|||||  
1656 GAGACCATTTTAAGATAAGATAACATCTCTCAGGCGACAGTACCATGAGCC 1705  
485 oLeuThrArgProSerGluGlnLeuAspTyrLeuSerArgValGlnGlyP 502  
|||||  
1706 TCTCAGAGACCCCTTGAGCAACTGACTATCTTTCAGAGATCCAGGAT 1755  
502 heGlnValGluTyrLysAspPheProLysAsnAsnLysAsnGluPheVal 518  
|||||  
1756 TCCAGGTGAATTAACAAAGACTTCCCAAAACAAACAAAGCAATTTTGA 1805  
519 SerLeuLLeAsnCysSerSerGlnProProLeuLLeSerHisGlyLLeG1 535  
|||||  
1806 TCTCTTATCAATTTGCTGCTCTGCAAGCCACTGATAGGCATGATGTCG 1855  
535 yLysAspValGluSerCysHisAspMetAlaAlaLeuAsnLLeuLysL 552  
|||||  
1856 CAAGGATGTGAGTCTCTGCCATGATGTGCTGGCGCAACATCTTAAAT 1905  
552 euLeuSerGluLeuAspGlnGlnSerThrGluLeuProArgThrGlyAsn 568  
|||||  
1906 TGTGTCTGATGTGAGCAACAAAGATGACAGATGTCACAGACAGGAAAC 1955  
569 GlyProMetSerValCysGlyArgCys 577  
|||||  
1956 GGACCAATGCTGCTGCTGCGGAGGTGC 1982  
seq\_name: gb\_pr4:AK025519  
seq\_documentation\_block:





```

1063 GGAAMCAGACTGACAGAGGACGGGACCACCAACAGAGGTGGCCAGCG 1112
342 GAENALALAGLUAASMETLEUGLUTLEUGLYPHELYSVALPROGLNA 359
1113 CATGACAGCCGAGACATGCTGGAGATCCTTGTTCAAGTCCCGAGG 1162
359 rglbProThrLysProAlaLeuLysSerGluLysThrProIleLys 375
1163 CGCAGGCCACCAACCCGACCTCAGATCAGAGAGAGAAGAACCCATTAAG 1212
376 LysProGlyAspGlyArgValThrPhePheGluProGlySerGlyAs 392
1213 AAACCCAGG: GATGAGAGAAAGTAACTTTTGAACCTGGCTCTGGGGA 1261
392 pgluAsnGlyThrSerAsnLysGluAspGluPheArgMetProTyrLeu 409
1262 TGAATAATGGAGTAAATTAAGAGATGAGTTGACAGATCCCTTAATCA 1311
409 eRhISGLInGluLeuProAlaGlyIleLeuProMetValProGluValAla 425
1312 GTGATCAGACAGTGCCTGCTGGAAATCTTCCCATGGTCCCGAGTCCGC 1361
426 GluAlaValGlyValSerGlnGlyHisIsthrLysAspPheThrArgAl 442
1362 CAGGCTGTAGAGTTAGTCAAGACATCACACCAAGATTTTACAGAGGC 1411
442 aaIaProAsnProAlaValAlaThrValThrAlaMetIleAlaArgGlu 459
1412 AGCTCCGAATCTGCCAGGCGCAGGTAACCTGCATGATAGCCCGAGAGT 1461
459 euLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysAsnAsn 475
1462 TGTGTATGGGGGACCTGCGCCACAGCGAGACCAATTTAAAGATTAAC 1511
476 IleSerSerGlyHisValProHisGlyProLeuThrArgProSerGlu 492
1512 ACCTCTTACAGGCGCCTACCCCATGGACCTCTCAGAGACCCCTCTGAGCA 1561
492 nueuAspTyrLeuSerArgValGlnGlyPheGlnValGluTyrLysAsp 509
1562 ACTGGACTATCTTCCAGAGTCCAGGGATTCAGGTTGAATACAAAGACC 1611
509 heProLysAsnAsnLysAsnGluPheValSerLeuIleAsnCySerSer 525
1612 TCCCAAAAACAACAGAGCAAGATTTGTATCTCTATCAATGCTCTCTCT 1661
526 GluProProLeuIleSerHisGlyIleGlyLysAspValGluSerCysH 542
1662 CAGCCACCTCTGATCAGCCATGATCGGCAGAGATGTGAGTCCCGCA 1711
542 sasPheTAlaIaLeuAsnIleLeuLysLeuLeuSerGluLeuAspGln 559
1712 TGATATGGCTGCGCTGACATCTTAAAGTTGCTGCTGAGTTGGACCAAC 1761
559 InseThrGluMetProArgThrGlyAsnGlyProMetSerValCysGly 575
1762 AAAGTACAGAGATGCCAAGACGAAAGCAACCATGTCTGTGTGG 1811
576 ArgCys 577
1812 AGGTGC 1817
seq_name: gb_pr6: HSA132258

```

```

seq_documentation_block:
LOCUS      HSA132258      3066 bp      mRNA      PRI      14-MAY-1999
DEFINITION Homo sapiens mRNA for stauflen protein, partial.
ACCESSION  AJ132258
VERSION    AJ132258.1 GI:4572587
KEYWORDS   stauflen gene; stauflen protein.
SOURCE      human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

```

```

REFERENCE 1 (bases 1 to 3066)
AUTHORS   Falcon,A.M., Fortes,P., Marion,R.M., Beloso,A. and Ortin,J.
TITLE      Interaction of Influenza virus NS1 protein and the human homologue
           of Stauflen in vivo and in vitro
JOURNAL    Nucleic Acids Res. 27 (11), 2241-2247 (1999)
MEDLINE    99263025
REFERENCE 2 (bases 1 to 3066)
AUTHORS   Ortin,J.
TITLE      Direct Submission
JOURNAL    Submitted (25-JAN-1999) Ortin J., Molecular and Cellular Biology,
           Centro Nacional de Biotecnologia, Cantoblanco, 28049 Madrid, SPAIN
FEATURES   Location/Qualifiers
            source          1..3066
                                /organism="Homo sapiens"
                                /db_xref="taxon:9606"
            gene            36..1526
                                /gene="stauflen"
            CDS              36..1526
                                /gene="stauflen"
                                /function="dsRNA binding protein"
                                /codon_start=1
                                /product="stauflen protein"
                                /protein_id="CAB40082.1"
                                /db_xref="GI:4572588"
                                /translation="MKLGKKPKPKVPDPSPSMOSTYNYNMKGAVPPRYEYPPVPL
                                LYQVELSVGGQFNGCKTQROAKHDAARLRLIQLQNPRLPERLVNGRSEFEENLAK
                                SEISOVEIALKRLNLPVNFVARSPPHKKNPTTKYSVGFVEGEGSKKIKKNA
                                AIAVLEELKLPPLPAVERVAPRIKKTKPLVKQTSPEYQGINPISRLQIQAKR
                                EKEPEYTLTERGLRRREYVQVGNHNAEGYTNKVAKRNAENMLELILGFVYP
                                QRSHTQYVAGEDTHKETRGMKSNLNLALGMENTSKNDEFRMPLISHQOLPA
                                GILMPVEVADAVSOGHHTKDTJRAAPNAPKATVYAMILARELLYGTSPATLTK
                                NISSGVHPGLTRPSQOLDYLSRVOGFQVEYKDPKNNKNEFVLSINSSQPLLS
                                HGICKVDESCDMAALNLIKLSLDQOSTMPRTGNGPMSVCGRC"
BASE COUNT      891 a      693 c      699 g      783 t
ORIGIN
alignment_scores:
Quality: 2577.50      Length: 510
Ratio: 5.104      Gaps: 4
Percent Similarity: 99.020      Percent Identity: 99.020
alignment_block:
US-09-316-048-6 x HSA132258
Align seg 1/1 to: HSA132258 from: 1 to: 3066
71 ILeThrProThrValGluLeuAsnAlaLeuCysMetLysLeuGlyLys 87
|||||
3 ATTAACCCCTAGCTGATGACTAAATGACATGTGATGAACTTGGAAAAA 52
87 sProMetTyrLysProValAspProTyrSerArgMetGlnSerThrTyr 104
|||||
53 ACCAATGATTAAGCCCTGTGACCCCTTACTCTCGGAGTGCACCTATA 102
|||||
104 sDTYrAsnMetIaArgGlyAlaTyrProProArgTyrPheTyrPrope 120
|||||
103 ACTTACACATGAGAGGAGGTGCTTATCCCGCAGAGTACTTTTACCATT 152
|||||
121 ProValProProLeuLeuTyrGlnValGluLeuSerValGlyGlnG 137
|||||
153 CCAGTTCACACCTTTACTTATCAAGTGAACCTTCTGTGGAGACAGCA 202
|||||
137 nPheAsnGlyLysGlyLysThrArgGlnAlaAlaLysHisAspAlaAla 154
|||||
203 ATTATATGCAAAAGCAAAAGACAGAGCTCGAAGAACATGCTGCTG 252
|||||
154 IalysAlaLeuArgIleLeuGlnAsnGluProLeuProGluArgLeu 170
|||||
253 CCAAACGTTTGAAGATCTGCAAGATGAGCCCTGCAAGAGAGGCTGAG 302
|||||
171 ValAsnGlyArgGluSerGluGluAsnLeuAsnLysSerGluIleSe 187

```

```

|||||
303 GTGAATGAGAGAAATCCGAAGAAATAATCTCAATAAATCTGAAATAG 352
187 rglValAlPhagluLeuAlaLeuLysArgAsnLeuProValAsnPhgIuV 204
353 TCAGATGTTGAGATGCTACCTTAAGCGAACTGCTGTGATTTCCAGG 402
204 aAlaArgGluSerGlyProProHisMetLysAsnPhaValThrLysVal 220
403 TGGCCGGGAGAGTGGCCCAACCACATGAAGAACTTTGTGACCAGATT 452
221 SerValGlyLupheValGlyGlyGlyGlyLysSerLysLysIleSe 237
453 TCGGTTGGGAGATTGTGGGGGAAGGTGAAGGCAAGCAAGAAATTC 502
237 rLysLysAsnAlaAlaIleAlaValLeuGluGluLeuLysLysLeuPro 254
503 AAGAAAAAATGCGCCCATAGCTGTGTTGAGGAGCTGAAGAAATTC 552
254 rLysProAlaValAlaGluArgValLysProArgLLeLysLysLysThrLys 270
553 CCGTCCCTGGAGTTGAACGAGTAAGCTTAGAATCAAAAAGAAACAAA 602
271 ProIleValLysProGluThrSerProGluTyrGlyGlnGlyIleAsnPr 287
603 CCGATAGTCAAGCCACAGCAAGACCCAGAAATATGCGAGGGGATCAATCC 652
287 oLleSerArgLeuAlaGlnIleGlnAlaIleLysLysGluLysGluProG 304
653 GATTAGCCGATGGCCAGATCCAGCAGGCAAAAAGAAAGAGAGAGCCAG 702
304 LuTyrThrLeuLeuThrGluArgLysLeuProArgArgArgGluPheVal 320
703 AGTACAGAGCTCTCAGAGAGAGAGCCTCCCGCCGAGGAGATTGTG 752
321 MetGlnValLysValGlyAsnHisThrAlaGluGlyThrGlyThrAsnLy 337
753 ATGAGAGTGAAGGTTGGAAACACACATGCAAGAAAGAGGCGCACCAAA 802
337 sLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyP 354
803 GAAGGTGGCCCAAGCGCAATGCAAGCCGGAACATCTGTGAGATCTCTGTT 852
354 heLysValProGluArgGluPro..ThrLysProAlaLeuLysSerGluG 370
853 TCAAGTCCCGCAGCGC...AGCCACCAAAACCCCACTCAAGTCAAGAGA 899
370 uLysThrProIleLysLysProGlyAspGlyArgLysValThr..PhePhe 386
900 GAAGACACCCATTAAGAAACCAAGGAGATGAAGAAAGTAACTTTT 949
387 GluProGlySerGlyAsp..GluAsnGlyThrSerAsnLysGluAspGluP 403
950 GAACCTGGCTCTGGGATGGAATAATGGAGTAAATAAGAGATAGT 999
403 heArgMetProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuPro 419
1000 TCAGGATGCCCTTATCTAAGTCATCAGCAGCTGCTGTGGAATCTTCC 1049
420 MetValProGluValAlaGlnAlaValAlaGlyValSerGlnGlyHisHis 436
1050 ATGGTGGCCGAGGTGCGCCAGGCTGTAGGATGTGTCAGACATCAACAC 1099
436 rLysAspPheThrArgAlaAlaProAsnProAlaLysAlaThrValThra 453
1100 CAAGATTTTACCGAGGCGACCTCCGAATCTGCGCAAGCGCCAGCTAATC 1149
453 lAmetIleAlaArgGluLeuLeuTyrGlyGlyThrSerProThrAlaGlu 469
1150 CCATGATAGCCCGAGAGATTGTTGATGGGGCACCTGCGCCACAGCCGAG 1199
470 ThrIleLeuLysAsnAsnIleSerSerGlyHisValProHisGlyProLe 486
|||||

```

```

1200 ACCATTTTAAAGATTAACATCTCTTACAGGCCACGTACCCCATGACCTCT 1249
486 uThrArgProSerGluGluLeuAspTyrLeuSerArgValGlnGlyPheG 503
1250 CACGAGACCCCTCTGACGACTGAGCTATCTTCCAGAGTCCAGGAGATTCC 1299
503 lValGluTyrLysAspPheProLysAsnAsnLysAsnGluPheValSer 519
1300 AGTTGATTAACAAGACTTCCCAAAAAACAAACAGAAAGAAATTTGTATCT 1349
520 LeuIleAsnCysSerSerGlnProProLeuIleSerHisGlyIleGlyLy 536
1350 CTTATCAATTGCTCTCTCACACCCTCTGTATCAGCATGGTATCGGCA 1399
536 sAspValGluSerCysHisAspMetAlaAlaLeuAsnIleLeuLysLeuL 553
1400 GGATGTGAGTCTCTGCCATGATATGCTGCTGCTGATCAATCTTAAAGTTGC 1449
553 euSerGluLeuAspGlnGlnSerThrGluMetProArgThrGlyAsnGly 569
1450 TGCTGAGTTGACCACAAAGTACAGAGATGCCAAGAAAGAAACGGA 1499
570 PrometSerValCysGlyArgCys 577
1500 CCAATGCTGTGTGTGGAGGTGC 1523

seq_name: gb_ro:AF290989

seq_documentation_block: 3041 bp mRNA 09-AUG-2000
LOCUS AF290989
DEFINITION Rattus norvegicus staufen isoform Stau-16 (Stau) mRNA, complete cds, alternatively spliced.
ACCESSION AF290989
VERSION AF290989.1 GI:9755162
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 3041)
AUTHORS Monshausen, M., Putz, U., Rehbein, M., Schweizer, M., Desgroselliers, L., Kuhl, D., Richter, D. and Kindler, S.
JOURNAL Two rat brain staufen isoforms differentially bind RNA
REFERENCE 2 (bases 1 to 3041)
AUTHORS Monshausen, M. and Kindler, S.
JOURNAL Direct Submision
SUBMITTED (28-JUL-2000) Institute for Cell Biochemistry and Clinical Neurobiology, University of Hamburg, University Hospital Eppendorf, Suderfeldstr. 24, Hamburg D-22529, Germany
FEATURES
source
1..3041
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="brain"
/dev_stage="adult"
1..3041
/gene="Stau"
227..1696
/gene="Stau"
/feature="note="RNA-binding protein; similar to Drosophila staufen; contains four RNA-binding domains; rStau-16"
/codon_start=1
/product="staufen isoform Stau-16"
/db_xref="GI:9755163"
/translation="MKYPDRHSRMOSTSYSGRGAYPRPYFPYPPVLLXVETLS VGGQFNGKGMKPPVKHDAATATLTLQSELPPELVNGBSEBENLMSKSEISQVF EIALKRNLPVNFVARESPGPMKNFVTVSVEFVGEDEGSKTSKRNAAAEVLEQ LRLPLPAVERVKPRIKKKSQPTCKLQTPYGGOMNPIISRLAQIQAKKREPEYM LLTERGLPRREFVMQVKVGHHTABGAGTNKKVAKRNAENMLEILGFKVPOQPAKP

```

ALKSEKTPYVKPGDGRKRVTFEESPQDENSTSNKDEFRMPYLSHOOLPAGILPMVY  
EVAQAVGSGSHHTKDFARAAAPNPAKATVTAMIRBELLYGTSPTAETILKSNTSGH  
VPHGPRTRPSOLYTLISRAQGFQVEYKDFPNKNMNECVSLINCSSQPTPLVSHGIGKDV  
ESCHDMAALNTLKLISELDQOSTEMPRGNGPVSAAGNC"  
polyA-signal  
/gene="Stau"

BASE COUNT 814 a 771 c 747 g 707 t 2 others  
ORIGIN

## alignment\_scores:

Quality: 2445.50 Length: 566  
Ratio: 4.712 Gaps: 4  
Percent Similarity: 91.696 Percent Identity: 84.452

## alignment\_block:

US-09-316-048-6 x AF290989 ..

Align seg 1/1 to: AF290989 from: 1 to: 3041

```

31 ProlumeSerIlePro.....SerThrth 39
   |||||::: |||
   7 CCCCTCTCTTGACCCCGCGCGCCGCGCTTCTCGCAGCCAC 56
   39 rSerSerLeuProSerGluAsn..... 46
   1 |||||::: |||
   57 TCCGTCCTTCCCTCCCGGCTTCTTCTCTGCGCCCTCTTGGCT 106
   47 ..AlaGlyArgProIleGlnAsnSerAlaLeuProSerAlaSerIleThr 62
   :::: ||| ||::: |||
   107 TTTCTCTCTCGCGCGCGCCCAATACCGCGCGCGG.....ACG 147
   63 SerThrSerAlaAlaAla..GluSerIleThrProThrValGluLeuAsn 79
   ||::: ||||| |||||::: ||::: |||||
   148 AGCTCTGGGGAGACACCCAGAAAGTATAGCTTCTACCATTTAGACTAATG 197
   79 lalaucysmethlysleuglylyslysprometryltyrsprovalaspro 95
   |||||::: ||||| :::: |||||::: |||||
   198 CACTGTGCGGAAATCGAAGAAACCAATGTATTAAGCCGTGAGACCC 247
   96 TyrSerArgMetGlnSerThrTyrAsnTyrAsnMetArgGlyGlyAlaTyr 112.
   :::: |||||::: |||||::: |||||::: |||||
   248 CACTCTCGGATGCACTCACCTACCTATGCATCGTGAGAGTCCATA 297
   112 rProProArgTyrPheTyrProPheProValProProlLeuTyrGlnY 129
   |||||::: |||||::: |||||::: |||||::: |||||
   298 TCCCCCAAGTACTTTTACCATTTCAGTTCCACCGTTACTTACCAAG 347
   129 alGluLeuSerValGlyGlyGlnGlnPheAsnGlyLysGlyLysThrArg 145
   |||||::: |||||::: |||||::: |||||::: |||||
   348 TTGAGCTCTCCGTGGGAGACAGCAGTTTAATGGAAAGGAAAGATGAGA 397
   146 GlnAlaAlaLysHisAspAlaAlaLysAlaLeuArgIleLeuGlnAs 162
   :::: |||||::: |||||::: |||||::: |||||
   398 CCACCTGTGAACATGATGCACCTGCCGCGAGCCCTGAGGACTCTGAGAG 447
   162 nGluProLeuProGluArgLeuGluValAsnGlyArgGluSerGluGluG 179
   :|||::: |||||::: |||||::: |||||::: |||||
   448 TGGAGCCCTTACAGAGAGCTTACAGTAATGGAAGAGAAATCAGAAAGAG 497
   179 lAsnLeuAsnLysSerGluIleSerGlnValPheGluIleAlaLeuLys 195
   |||||::: |||||::: |||||::: |||||::: |||||
   498 AAAACCTTATAATCAGAAATTAAGCAAGTGTGTGAGATTGGCGTGAAG 547
   196 ArgAsnLeuProValAsnPheGluValAlaArgGluSerGlyProProH 212
   |||||::: |||||::: |||||::: |||||::: |||||
   548 CGCAATTTGCTGTGAATTTTGGAGTGCGCCGCGAGAGTGCGCCGACACA 597
   212 smetLysAsnPheValThrLysValSerValGlyGluPheValGlyGluG 229
   |||||::: |||||::: |||||::: |||||::: |||||
   598 CATGAAGAACCTTGTGACCAAGGCTTACAGTTGGGAGTTGTGAGGGAGAG 647
   229 lylGluLysSerLysLysIleSerLysAsnAlaAlaIleAlaVal 245

```

```

648 GAGAAAGGAAAGCAAGAGATCTCCAGAAAGATCGGCAAGGCGCGTG 697
246 LeuGluGluLeuLysLysLeuProProlLeuProAlaValAlaGluArgValY 262
|||||::: |||||::: |||||::: |||||::: |||||
698 CTGGAGCACTTAAGAGGCTGCGCACCTTACCTGCTGTGGAGGAGTGA 747
262 sProArgIleLysLysLysThrLysProIleValLysProGlnThrSerP 279
|||||::: |||||::: |||||::: |||||::: |||||
748 GCCCAGATCAGAGAGAGAGAGTACGCCCTGACAGCTACAGACTGCC 797
279 roGluTyrGlyGlnGlyIleAsnProIleSerArgLeuAlaGlnIleGln 295
||::: |||||::: |||||::: |||||::: |||||
798 CCGACTATGGCCAAAGGATGAATCTATTAGCAGACTGCGACAGATCCAG 847
296 GlnAlaLysLysGluLysGluProGluTyrThrLeuLeuThrGluArgG 312
|||||::: |||||::: |||||::: |||||::: |||||
848 CAGCAAAAAAGAGAGAGAGCGGAGTACATGCTCTTACAGAACGAGG 897
312 lLeuProArgArgArgGluPheValMetGlnValLysValGlyAsnHisT 329
|||||::: |||||::: |||||::: |||||::: |||||
898 TCTTCCACGTCGAGGAGATTGTGTGATGACAGTGAAGTTGGGCATCACA 947
329 hralaGluGlyThrGlyThrAsnLysLysValAlaLysArgAsnAlaAla 345
|||||::: |||||::: |||||::: |||||::: |||||
948 CCGCAGAGAGAGCAGGTACCAATTAAGAGTGCGCAAGCGCAATGCTGCT 997
346 GluAsnMetLeuGluIleLeuGlyPheLysValProGlnArgGlnProH 362
|||||::: |||||::: |||||::: |||||::: |||||
998 GAGAACATCTGGAGATCTGTGGGTTCAAGTCCCGCAGCGGAGGCTCG 1047
362 rLysProAlaLeuLysSerGluGluLysThrProIleLysLysProGlyA 379
|||||::: |||||::: |||||::: |||||::: |||||
1048 CAAGCCAGCACTCAATTCAGAGAGAGAGACTCCAGTAAGAAACAGAGAG 1097
379 spGlyArgLysValThrPhePheGluProGlySerGlyAspGluAsnGly 395
|||||::: |||||::: |||||::: |||||::: |||||
1098 ATGGAAGAGAAAGTACATTTTGAACCTAGCCCTGCGGATGTAAATGGA 1147
396 ThrSerAsnLysGluAspGluPheArgMetProTyrLeuSerHisGlnG 412
|||||::: |||||::: |||||::: |||||::: |||||
1148 ACCAGTAATTAAGAGATGATGTCAGATGCCCTATCTTAAGCCATAGCA 1197
412 nLeuProAlaGlyIleLeuProMetValProGluValAlaGlnAlaValG 429
|||||::: |||||::: |||||::: |||||::: |||||
1198 GCTGACAGCGCGGATTTCTCCCATGTGCGCAGAGTTGCCCAAGCTGTGG 1247
429 lYValSerGlnGlyHisHisThrLysAspPheThrArgAlaAlaProAsn 445
|||||::: |||||::: |||||::: |||||::: |||||
1248 GGGTTAGTCAAGAGCACCATACCAAGATTTCCCGAGGCGAGCTCCAAAT 1297
446 ProAlaLysAlaThrValThrAlaMetIleAlaArgGluLeuLeuTyrG 462
|||||::: |||||::: |||||::: |||||::: |||||
1298 CCTGCCAAGGCGCGGTACTGCTCATGATAGCCCGGAATTTGTTAGCGG 1347
462 yGlyThrSerProThrAlaGluThrIleLeuLysAsnAsnIleSerSerG 479
|||||::: |||||::: |||||::: |||||::: |||||
1348 GGGCACCCTGGCCACAGCGGAGACATTTTAAGAGTATACATCTTTCAG 1397
479 lYHisValProHisGlyProLeuThrArgProSerGluGlnLeuAspTyr 495
|||||::: |||||::: |||||::: |||||::: |||||
1398 GACATGTACCCCATGAGCCTCGCACAAAGCCCTCAGCAACATGTCCTAC 1447
496 LeuSerArgValGlnGlyPheGlnValGluTyrLysAspPheProLysAs 512
|||||::: |||||::: |||||::: |||||::: |||||
1448 CTTTCCAGAGCCCGAGGATTCAGGTTGAATTAAGAACTCTCCCAAGAA 1497
512 nAsnLysAsnGluPheValSerLeuIleAsnCySerSerGlnProProL 529
|||||::: |||||::: |||||::: |||||::: |||||
1498 CAACAAGAACAGACTGTATCTCTCATCAACTGCTTTCACAGCCACCTC 1547
529 euIleSerHisGlyIleGlyLysAspValGluSerCysHisAspMetAla 545
||::: |||||::: |||||::: |||||::: |||||

```

```
1548 TCGTCACCATGCGATCGGACAGATGTGAGTCCTGTCATGATATAGGCT 1597
546 AlaleuansnileuleuLysleuleuSerGluLeuaspGlnSerThrG1 562
|||||
1598 GCACTGAACATCTTAAGCTCTCTGTGAGTTGACCAAGACAGACAGCA 1647
562 uMetProArgrThrGlyasnGlyPrometSerValCysGlyArgCys 577
|||||
1648 GATGCCAAGACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1693
seq_name: gb_ro:AF227200

seq_documentation_block:
LOCUS AF227200 3059 bp mRNA 09-AUG-2000
DEFINITION Rattus norvegicus staufen isoform Stau+16 (Stau) mRNA, complete
cds, alternatively spliced.
ACCESSION AF227200
VERSION AF227200.1 GI:9754865
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 3059)
REFERENCE Monshausen, M., Putz, U., Rehbein, M., Schweizer, M.,
AUTHORS Desgroselliers, L., Kuhl, D., Richter, D. and Kindler, S.
TITLE Two rat brain staufen isoforms differentially bind RNA
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3059)
AUTHORS Monshausen, M. and Kindler, S.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2000) University of Hamburg, Institute for Cell
Biochemistry and Clinical Neurobiology, Martinistrasse 52, Hamburg
20246, Germany
FEATURES
source
1.3059
Location/Qualifiers
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/cisue_type="brain"
/dev_stage="adult"
1.3059
/gene="Stau"
227..1714
/gene="Stau"
/feature="RNA-binding protein; similar to Drosophila staufen;
contains four RNA-binding domains; rStau+16"
/codon_start=1
/product="staufen isoform Stau+16"
/protein_id="AAF98119.1"
/db_xref="GI:9754866"
/translation="MTKPDVPHSRMSTYSYGNRGAGVPPRYFPVPPPLLYOVELS
VGQDFNGKMKPPVAKHDATRALRTLOSEPRLPELVNGRSEBENLKSISQVF
EIALKRLNLEVNFESEPLTOVARESGPAMKNFTYRVSGVEFGESGSKISKNA
RAVLBOLRLPLPAVERVPRPKKSOPTKLOTAPDVGOGNNPISRLAIOQAKKE
KPEVVALTERGLPRRFRFVMOYKVGHTAAGTAKKYAKNMLETIGFKVPQ
APPAKAKLSEKTPYKKRGDGGKTYTFPERSPDEGTSKREDEFMPYISHOQLAG
ILPMPEVAQAVGVSQGHHTKDPARAAPAKATVAMLAIRELLYGTSPTAETILKS
NISSGVHPGPTRPSEOLYILSRAGFOVEYDPRKNNKNECVSLINCSOPPLVSH
GIKGVESCHDMAALNLIKLSLSDOSTEMPTGNGPVSACGRC"
polya_signal
3043..3048
/gene="Stau"
BASE COUNT 817 a 778 c 749 g 713 t 2 others
ORIGIN
alignment_scores:
Quality: 2432.50 Length: 572
Ratio: 4.687 Gaps: 5
Percent Similarity: 90.734 Percent Identity: 83.566
alignment_block:
```

```
US-09-316-048-6 x AF227200 ..
Align seg 1/1 to: AF227200 from: 1 to: 3059
31 ProleuMetSerIlePro.....SerThrTh 39
|||||
7 CCGCTTCTTCGACCCCGGCGCGCGCGCGCGCTTCTCGCAGCCAG 56
rSerSerIleProSerGluasn..... 46
1 TCGGCTCTTCCCTCCCGCGCTTCTTCTCGCCTCTTCTTCTTCT 106
47 ..AlaGlyArgProIleGlnAsnSerAlaLeuProSerAlaSerIleThr 62
|||||
107 TTTCTTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 147
SerThrSerAlaAlaAlaGluSerIleThrProThrValGluLeuAsn 79
|||||
148 AGCTCTGGGGAAGCAGCCAGCAAGATAGCTTCAACATGAGCTCAATG 197
79 lAleuCysMetIleuSerGlyLysLysProMetIleProValAspPro 95
|||||
198 CACTGTGCTGAACTGGAAAGAAACCAATGATATAGCCGTGACCC 247
TyrSerArgMetGlnSerThrTyrAsnTyrAsnMetArgGlyGlyAlaTy 112
|||||
248 CACTGTGCTGAACTGGAAAGAAACCAATGATATAGCCGTGACCC 297
ProProArgrTyrPheTyrProPheProValProProLeuLeuTyrGlnV 129
|||||
298 TCCTCCCGAGATACCTTTCACCATTCAGTTCACCGTTCACCTACCAAG 347
129 alGluLeuSerValGlyGlyGlnGlnPheAsnGlyGlyGlyThrArg 145
|||||
348 TTGAGCTCTCCGTGGAGAGACGAGCTTAAAGGAAAGGAAAGTGA 397
GlnAlaAlaLysHisAspAlaAlaAlaLysAlaLeuAlaGlyLeuGlnAs 162
|||||
398 CCACCTGTGAACATGATGCCATCCCGCGAGCCGTGAGACTCTGACAG 447
CGAGAGTGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 597
62 gClnSerGlyProProHisMetLysAsnPheValThrLysValSerValG 223
598 GGAGAGTGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 647
223 lGluLpPheValGlyGlnGlyGlnGlyLysSerLysLysIleSerLys 239
648 GGAATTTGTAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 697
240 AsnAlaAlaIleAlaValLeuGlnGluLeuLysLysLeuProProLeuPr 256
|||||
698 AATGGGCGAAGGCGCGCTGTGAGCAGCTAAGAGAGCGCCACCTCTAC 747
748 TGCTGTGAGCGAGTGAAGCCAGATCAAGAAAGATCAGCCACCT 797
256 alAlaValGluArgValLysProArgIleLysLysLysThrLysProIleV 273
|||||
273 allYsProGlnThrSerProGluTyrGlyGlnGlyIleAsnProIleSer 289
|||||
798 GCAAGCTTACAGACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 847
290 ArgLeuAlaGlnIleGlnAlaLysLysGlnLysGlnProGluTyrTh 306
```

```

848 AGACTGGCACAGATCCAGACGCGCAAAAAAGAGAGAGACCGCGGAGTACAT 897
|||||
306 rleuleuThrGluArgGlyLeuProArGArGArgGluPheValMetGlnv 323
|||||
898 GCTCTTACAGAACGAGGCTTTCACCTCCAGGAGATCTTGATGTCAGG 947
|||||
323 allysValGlyAsnAsnIstThrAlaGlnGlyThrGlyThrAsnLysVal 339
|||||
948 TGAAGGTGGGGATCTACACCGCGAGAGGACAGGTACCAATTAAGAAAGTG 997
|||||
340 AlaLysArgAsnAlaAlaGluAsnMetLeuGluLeuGluPheLysVal 356
|||||
998 GCCAAGCGCAATGCTGCTGAGAACATGCTGAGATCTCGGGGTTCAAGT 1047
|||||
356 lProGlnArgGlnProThrLysProAlaLeuLysSerGluLysThrP 373
|||||
1048 CCCCAGGCGCCACCTGCGCAAGCCAGCATCTCAATCAGAAAGAAAGCTC 1097
|||||
373 rOleLysLysProGlyAspGlyArgLysValThrPhePheGlnProGly 389
|||||
1098 CAGTAAAGAAACACAGAGATGAGAGAAATACATTTTGTGAACCTTAGC 1147
|||||
390 SerGlyAspSyluAsnGlyThrSerAsnLysGluAspLupPheArgMetPr 406
|||||
1148 CCTGGGATGAATAATGGAACACAGTAATTAAGAGATGATGTTACAGATGCC 1197
|||||
406 oTyrrLeuSerHisGlnGlnLeuProAlaGlyIleLeuProMetValProG 423
|||||
1198 CTATCTTAGGCATCAGACAGCTGCCAGCGGGATTTCTCCATGGTGGCAG 1247
|||||
423 luValAlaGlnAlaValGlyValSerGlnGlyHisIstThrLysAspPhe 439
|||||
1248 AGGTGGCCACGAGCTGTGGGGTTAGTCAAGACACCATACCAAAAGATTTC 1297
|||||
440 ThrArgAlaAlaProAsnProAlaLysAlaThrValThrAlaMetIleAl 456
|||||
1298 GCCAGGCGACACTCCAAATCTGCCAAGGCCACCGTAACTCCATGATAGTC 1347
|||||
456 aaArgGluLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeu 473
|||||
1348 CCGAGATTTGTTTACGGGGGACCTGCCACAGCCGAGAACCATTTTAA 1397
|||||
473 yAsnAsnIleSerSerGlyHisValProHisGlyProLeuThrArgPro 489
|||||
1398 AGAGTAAACATCTCTTCAGACATGTACCCCATGAGCTCCGACAAAGACC 1447
|||||
490 SerGluGlnLeuAspTyrLeuSerArgValGlnGlyPheGlnValGluTy 506
|||||
1448 TCTGAGCAACTGTACTACTCTTCCAGAGCCACGAGATTCCACAGTTGATA 1497
|||||
506 rLysAspPheProLysAsnAsnLysAsnGluPheValSerLeuIleAsn 523
|||||
1498 CAAAGACTCTCCCAAGAACAAAGAAAGCGAGTGTATCTCTATCACTACT 1547
|||||
523 ySerSerGlnProProLeuIleSerHisGlyIleGlyLysAspValGlu 539
|||||
1548 GCCTCTACAGCCACCTCTGCTGACGACCATGGCATCGGCAAGAGTGGAG 1597
|||||
540 SerCysHisAspMetAlaAlaLeuAsnIleLeuLysLeuLeuSerGlu 556
|||||
1598 TCTGTGCATGATATGCTGCTGACGACCATCTTAAAGCTGCTGTCTGAGTT 1647
|||||
556 uaSerGlnGlnSerThrGluMetProArGThrGlyAsnGlyProMetSerV 573
|||||
1648 GGACCAACACAGACAGAGATGCCAAGGACGAGAAAGCAAGCAAGTTTCAG 1697
|||||
573 alCysGlyArgCys 577
:::|||||
1698 CGTGCGGAGGTGC 1711

```

```

seq.documental_block: 2860 bp mRNA ROD 04-MAR-1999
LOCUS AF061942 Mus musculus RNA-binding protein staufen mRNA, complete cds.
DEFINITION Mus musculus RNA-binding protein staufen mRNA, complete cds.
ACCESSION AF061942
VERSION AF061942.1 GI:4335944
KEYWORDS
SOURCE
ORGANISM
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 2860)
Wichman,L., Duchaine,T., Luo,M., Nabl,I.R. and Desgroselliers,L.
Mammalian staufen is a double-stranded RNA- and tubulin-binding
protein which localizes to the rough endoplasmic reticulum
Mol. Cell Biol. 19 (3), 2220-2230 (1999)
JOURNAL
9917057
MEDLINE
2 (bases 1 to 2860)
Duchaine,T., Luo,M. and Desgroselliers,L.
AUTHORS
Direct Submission
JOURNAL
Submitted (28-APR-1998) Biochemistry, University of Montreal, 2900A
Edouard Montpetit, Montreal, Que H3T 1J4, Canada
FEATURES
source
1..2860
/organism="Mus musculus"
/db_xref="taxon:10090"
326..1789
/codon_start=1
/product="RNA-binding protein staufen"
/protein_id="AADI7529.1"
/db_xref="GI:4335945"
/translation="MYKVPDPHPSRMQSTYSYGMRGAGYPPRYFPYPPVPLLYQVEL
VGOQOFNGKGRPVKHDAPARALTTLOSEPRLPELVNGREAEENLNKSEISQV
EALRLNLPVNEVARRESGPPMKNNVTTSVGEFVGEEGSKISKKNAAVALE
LRRLPLPAVERVRKPRIKKKSQPTCTADYDGGNMPISRLAIOAKKEPEVMV
TERGCPRRREFYMWQKVYGHHTAEGCTGNKNKVAKRNAEMMLTELGTGYQAQAPKPR
KSEETTPVKKPGDGRKVTVEFEPSPGNETSKNDEFNRPYLSHQLPGLPWPE
AGAVSGOHHKRDFTRAAPNPAKATVAMKALVELLYGTSPTAETILNSISSGHV
HQPRTPSBOLLYLSRAQGFQVEYKDFPNMNNNECVSLINCSSOPLVSHIGIKDVE"
CHDMAINLTIKLSELDQOSTEMPRNGNPVSAACGRG"
BASE COUNT 747 a 750 c 713 g 650 t
ORIGIN
alignment_scores:
Quality: 2422.50 Length: 552
Ratio: 4.750 Gaps: 2
Percent Similarity: 92.391 Percent Identity: 85.145
alignment_block:
US-09-316-048-6 x AF061942 ..
Align seg 1/1 to: AF061942 from: 1 to: 2860
26 SerLeuLeuSerGlnProLeuMetSerIleProSerThrThrSerSerIle 42
||||| ::| ::| ::| ::| ::| ::| ::| ::|
158 TCCCTCCTCGTGGCGGCCCTTTTCTCCGCCGCTTACACTGCTTTC 207
42 uProSerGluAsnAlaGlyArgProIleGlnAsnSerAlaLeuProSerA 59
|||| ::| ::| ::| ::| ::| ::| ::| ::|
208 ACCTCCTCGCGGCCGCCAAGA.....CCGGCCG 236
59 IaSerIleThrSerThrSerAlaAlaIaGluSerIleThrProThrVal 75
|| ::| ::| ::| ::| ::| ::| ::| ::|
237 CCCCCGACGAGCTCTGGGAGAGCAGCAAGAAATATAGCTTACCAT 286
76 GluLeuAsnAlaLeuGlyMetLeuGlyLeuGlyLysProMetTyrLysPr 92
||||| ::| ::| ::| ::| ::| ::| ::| ::|
287 GAGCTCAATGACACTGTGTGAACACTGGAAGAAACCAATGATTAAGCC 336
92 oValaIaProTyrSerArgMetGlnSerThrTyrAsnTyrAsnMetArg 109
||||| ::| ::| ::| ::| ::| ::| ::| ::|
337 CGTGGAGCCCTCTCTCGATGACAGTCCACCTAAGCTATGACGTGGTGG 386

```

109 *lyg1*ValaLysProProAlaGlyThrPheLysProPheProValProLeuVal 125  
106 *lyg1*ValaLysProProAlaGlyThrPheLysProPheProValProLeuVal 125  
387 GAGGGCCCATCCCCCAGATACCTTTTACCATTCCACGTCCACCTTTA 436  
126 *Leu*YrGLInValGlnLeuSerValGlyGlnGlnPheAsnGlyLysGln 142  
437 CTTCCACCAAGTTGAGCTCCGTGGGCGGACAGCAAGTTTAAATGGAAAG 486  
142 *lyl*ThrArGlnAlaAlaLysHisAspAlaAlaLysAlaLeuArGln 159  
487 AAGAGTGAACACCCCGTGAACACGATGCCCTCCGTGGCGTGAGGA 536  
159 *Leu*GLInAsnGlnProLeuProGluArGlnGlnValAsnGlyArGln 175  
537 CTTCTGAGGTGAACCCCGCCAGAAAGTTGGAGTAAATGGAAGACAA 586  
176 *Ser*GlnGlnGlnAsnLeuAsnLysSerGluLysSerGlnValPheGlnL 192  
587 GCAGAGAGAAAGAAACCTCATATATCGCAATTAACCAAGCTTTGAAAT 636  
192 *Leu*AlaLeuLysArGAsnLeuProValAsnPheGlnValAlaArGlnL 209  
637 TGGCGTGAAGCGGAATTTCCTGTGATTTTGAAGTGGCGCGGAGACTG 686  
209 *Lys*ProPheHisMetLysAsnPheValThrLysValSerValGlyLysPhe 225  
687 GCCCACACACATGAAAGAACTTTGTACCAAGGGTTTCAGTTGGGGAATTT 736  
226 *Val*GlnGlnGlnGlnLysSerLysLysLysLysLysLysLysAsnAlaAla 242  
737 GTTAGGGGAAGGAGAAAGGAAAGCAAGAAAGATCTCCAAAGAAAGATCG 786  
242 *Ala*LeuAlaLysGlnGlnGlnLysLysLysLysLysLysProLeuProAlaValG 259  
787 CAGGGCTGTTCTGAGACACTTAAGGAGGTGCCACCCCTCCCTGCTGAG 836  
259 *Lys*ArgValLysProArGlnLysLysLysLysThrLysProLysValLysPro 275  
837 AGCGATGTGAAGCCCAAGAAATCAAGAAAGAAAGTCCACCCACCTGCAAG 883  
276 *Gln*ThrSerProGlnLysLysLysLysLysLysLysLysLysLysLysLys 292  
884 ...ACAGCCCCGAGATTATGGCCAGAGGATGAATCTATTATGAGCACTTGC 930  
292 *Asn*LysGlnGlnAlaLysLysGlnLysGlnLysGlnLysProGlnLysLysLys 309  
931 ACAGATCCAGCAGCAAAAGAGAGAGAGCAGCAAGTACATGCTCCTTA 980  
309 *His*GlnArgGlyLeuProArGArGArGlnPheValMetGlnValLysVal 325  
981 CAGAACGAGAGTCTTCCACGTCCAGAGGAGTTGTATCAGGTAAAGGTT 10330  
326 *Gly*AsnHisThrAlaGlnGlnLysHisLysHisLysValAlaLysArG 342  
1031 GGGCATACACATGCGCAAGAGAGATGGGTACCAATTAAGAAAGGTGCCAACG 10800  
342 *Asn*AlaAlaGlnLysAsnMetLysGlnLeuGlnLysLysValProGlnAla 359  
1081 TATAGTCTGTGAGAACATCTGTGAGATCTCGGGGTTCAAAGTTCCCCAG 11330  
359 *Arg*GlnProThrLysProAlaLeuLysSerGlnGlnLysThrProLysLys 375  
1131 CCGAGCTCCCAAGCCAGCAGCATCAATCAGAAAGAGACCTCCAGTAAAG 11800  
376 *Lys*ProGlnLysArgGlyArgGlyValThrPhePheGlnLysProGlnLys 392  
1181 AACCCGAGAGAGCGAAGGAAATTAACGTTTTTTGAACCTAGCCCTGGGGA 12300  
392 *Arg*LysAsnGlyThrSerAsnLysGlnAlaAspGlnPheArgMetProLysLys 409  
1231 TGCAATATGGAAGTACAGTACACAGACAGAGAGTTCAGAGTGCCTATCTTA 12800  
409 *Thr*HisGlnGlnLeuProAlaGlyLysLeuProMetValProGlnValAla 425

```

/codon_start=1
/product="39k3 protein"
/protein_id="CAB64341.1"
/db_xref="GI:6534457"
/translation="MDIINOMFSVQLSLGFTWSESSSTKKQAQVANKALESTLP
KPVOKPKSNVNNPNSGITPTVELNGLAMKRGEEALTRLDKPFPPVRYRANVFRKAT
NORHCHVPKPIFYVQLTVGNNEFEGRKQAAHNAALALALONEPIPERSPONG
ESGKMDDDKANKSEISLVEIETALKRMKPEVSEIPEKSPPHKSFVTVSVGEFSA
EGGNSKKLSKKRAATVLOELKPLPVKEKPLFEKKPKTIYKAGEYGGGNP
ISRLAIOQAKKKEKPEPVYLSERGMRRRPFVQVGNVEATGTGPNKKIAKKNA
EAMILQGTAKSTRMLDQLEKTEGKNSGPKRPEPTNNTPRGIIHLSPDYQEME
ASRHKVISGTTGLTSLPKDMNPSSSEFSSISPTNSSATLTARELLMNGTSTVAICL
KSSPPPCSPVPSKQLEVLARIQGFV"

BASE COUNT      1280 a      731 c      737 g      1310 t
ORIGIN

alignment_scores:
    Quality: 1081.50      Length: 539
    Ratio: 2.831      Gaps: 19
    Percent Similarity: 70.872      Percent Identity: 48.794

alignment_block:
US-09-316-048-6 x HSA9062 ..

Align seg 1/1 to: HSA9062 from: 1 to: 4058

1 MetSerGlnValGlnValGlnValGlnAsnProSerAlaAlaLeuSerG1 17
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
220 AGTTCTTCAGTCAGCTGAGTCTTGTTGAGCAGACATGGGAATCCGAAAG 269
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17 YserGlnIleuAsnLysAsnGlnSerLeuLeuSerGlnProLeuMetS 34
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
270 CAGCAGTAAAGAAAGGCTCAGCAGGCTCTGCCAATAAGCTTTC... 315
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
34 ex1LeProSerThrThrSerSerLeuProSerGlnAsnAlaGlyArgP 50
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
316 .....ACTGAATCTACGCTTCC.....AAACA 339
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
51 IleGlnAsnSerAlaLeuProSerAlaSerIleThrSerThSerAlaAl 67
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
340 GTTCAGAACCCA.....CCCAAAAGTAATGTTAACAAATACCAAGGC.. 381
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
67 alaGlnSerIleThrProThrValGluLeuAsnAlaLeuGlyMetLysL 84
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
382 .....AGTATACTCCAACTGTGGAAGTGAATGGCTTGCCTATGAAA 424
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
84 euGlyLysLysProMetLysLysProValAspProLysSerArgMetGln 100
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
425 GGGGAGAGCGCTGCCATCTACAGCCATTGATCCAAAGCCATTCCCAAT 474
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
101 SerThrTyraThrAsnMetArgGlyGlyAlaTyrrProProArgTyrr 117
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
475 TATAGAGCTAATTAACAATTTCCG...GGCATGTACAATCAGAGGTATCA 521
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
117 eTyrrProPheProValProProLeuLeuTyrrGlnValGluLeuSerVal 134
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
522 T.....TGCCAGTGCCTTAAGATCTTTAT...GTTCAGCTCAGCTGAG 562
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
134 IyGlyGlnGlnPheAsnGlyLysGlyLysThrArgGlnAlaAlaLysHis 150
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
563 GAAATAATGAATTTTGGGAAAGAAAGACTGACAAAGCTGCTAGACAC 612
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
151 AspAlaAlaAlaLysAlaLeuArgIleLeuGlnAsnGluProLeuProG1 167
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
613 AATGTGTCAGAAATGAAGCCCTCCAGCAGCAAGTAAGAACCTATTCCACA 662
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
167 uArgLeuGlnValAsnGlyArgGlu.....SerGluGluGluA 180
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
663 AAGATCTCTCAGATGCGTAATCAGAAAGCATATGATGATGACAAAG 712
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
180 snLeuAsnLysSerGluIleSerGlnValPheGluIleAlaLeuLysArg 196
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

713 ANGCAAATAGTCTGAGATCAGCTAGTGTGTAATTCCTCTGAGACGA 762
197 AsnLeuProValAsnPheGluValAlaArgGluSerGlyProProHisMe 213
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
763 AATATGCCCTGTCAGTTTGTAGGTTATTAAAGAAAGTGACCAACCATAT 812
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
213 LlysAsnPheValThrLysValSerValGlyGluPheValGlyGlyG 230
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
813 GAAAGACTTTGTACTCAGAGTGCAGTAGAGAGATGCTCTGCAAGAGAG 862
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
230 IyGlyLysSerLysLysLysSerLysLysAsnAlaAlaAlaLeuLeu 246
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
863 AAGGAATAGCAAAAACCTCCAAAGAGCGCGTCCGACCAACCGCTTAA 912
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
247 GluGluLeuLysLysLeuProProLeuProAlaValGluArgValLysPr 263
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
913 CAGGAGCTTAAAAAACCTTCACCTCTTCTGTGTGTGGAAGCAACCAACT 962
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
263 oArgIleLysLysLysThrLysProIleValLysProGlnThrSerProG 280
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
963 AATTTTAAAAAAGCGCCATAAACAATAGTAAG.....GCCGAGCAG 1006
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
280 IuTyrrGlyGlnGlyIleAsnProIleSerArgLeuAlaGlnIleGlnG 296
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1007 AATATGCCCAAGGATGAACCTATTAGCCGCTGCGCAAAATTCACAG 1056
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
297 AlaLysLysGluLysGluProGluTyrrThrLeuLeuThrGluArgGlyL 313
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1057 GCCAAAGAGAAAGAGCGCGATATGTTTGTCTTTCGAAAGAGGAT 1106
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
313 uProArgArgArgGluPheValMetGlnValLysValGlyAsnHisThra 330
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1107 GCCTGAGCTCGAGAAATTTGTATGAGGTGAAGTAGCAATGAAGTTG 1156
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
330 IaGluGlyThrGlyThrAsnLysLysValAlaLysArgAsnAlaGlu 346
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1157 CTACAGCAAGACGACCTAATAAAGATGACCAAAAATAATCTCGAGAA 1206
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
347 AsnMetLeuGluIleLeuGlyPheLysValProGlnArgGlnProThrLy 363
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1207 GCATATGCTGTTCAACCTGTGTTATTAAGCATCCACTAAT..... 1245
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
363 sProAlaLeuLysSerGlu...GluLysThrProIleLysLysProGlyA 379
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1246 .....CTTCAGATCACTGTGAGAGACAGGAGAAACAAAGATGGA 1288
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
379 sPGlyArgLysValThrPhePheGluProGlySerGlyAspLysGlnGly 395
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1289 GTGGTCCAAAGCCTGGGTTTCTCTGAACCA..... 1317
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
396 ThrSerAsnLysGluAspGluPheArgMetProTyrrLeuSerHisGlnG 412
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1318 ACNAATAAAT..... 1326
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
412 nLeuProAlaGlyIleLeuProMetValProGluValAlaGlnAlaVal 428
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1327 .ACTCCAAAGAAATTCCTCATTTGTCCTCATGTTTCAAGAAATGG 1375
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
429 .....GlyValSerGlnGlnHis...His 435
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1376 AAGCAGCGCCGCAAAAGTAATCTCGGCATACCTAGGCTATTTGTCA 1425
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
436 ThrLysAspPheThrArgAlaAlaProAsn.....ProAlaLy 448
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1426 CCCAAGATATACACCACTTCAGCTTTTCTTCAGATATCTCCAC 1475
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
448 saIaThrValThrAlaMetIleAlaArgGluLeuLeuTyrrGlyGlyHis 465
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1476 ATCGAATAGTTCAGCTACAAATGCCAGCACTCTTATGAATGGAACAT 1525
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
465 eTyrrProThrAlaGluThrIle...LeuLysAsnAsnLysSerGlnHis 480
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1526 CTTCATACAGCTGAAGCCTAGGTTTAAAGGA.....AGTCTCTCTACT 1569
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```



```

481 ValProHisGlyProLeuThrArgProSerGluInLeuAspTyrLeuSe 497
      ||| ::::::::::: ::::::::::: ||::::::::: ||::
1570 CCCCTTGTTCTCCAGTA...CAACCTTCAAAACACTGGATATTATGAC 1616
      ||| ::::::::::: ||::::::::: ||::::::::: ||::
497 rArgValGlnGlyPheGlnValGluTyrLysAspPheProLysAsnAsnL 514
      ||| ::::::::::: ||::::::::: ||::::::::: ||::
1617 AAGGATTCAAGGCTTTCAGGTGATGAATTAAGCAAAACAAAAACAAA 1666
      ||| ::::::::::: ||::::::::: ||::::::::: ||::
514 yAsnGlnPheValSer 519
      ::::::::::: |||
1667 ACAATTCATTAGCCTCA 1683

seq_name: gb-p14:AK002152

seq_documentation_block:
LOCUS AK002152 2968 bp mRNA PRI 22-FEB-2000
DEFINITION Homo sapiens cDNA FL111290 fis, clone PLACE1009622, weakly similar
to MATERNAL EFFECT PROTEIN STAUEN.
ACCESSION AK002152
VERSION AK002152.1 GI:7023857
KEYWORDS oligo capping: fis (full insert sequence).
SOURCE Homo sapiens placenta cDNA to mRNA, clone_11b:PLACE1
clone:PLACE1009622.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
Wagatsuma,M., Hosokiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
Ono,Y., Takiguchi,S., Matanabe,S., Kimura,K., Murakami,K.,
Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
Nakamura,Y., Nagahari,K., Masuno,Y., Ninomiya,K. and Iwayanagi,T.
NEDO human cDNA sequencing project
Unpublished (2000)
2 (bases 1 to 2968)
Isogai,T. and Otsuki,T.
DIRECT SUBMISSION
Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases. Takao
Isogai, Helix Research Institute, Genomics Laboratory; 1512-3 Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
Tel:81-438-52-3951, Fax:81-438-52-3952)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan, cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing and clone selection;
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.

FEATURES
source
location/Qualifiers
1..2968
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PLACE1009622"
/clone_11b="PLACE1"
/tissue_type="placenta"
/Note="Cloning vector: pME18SFL3"
255..1967
/Note="unnamed protein product"
/codon_start=1
/db_xref="GI:7023858"
/translation="MANPEKTRAMCLVNELARNRVOYQKILNKGPAHSKMFVSOL
SLGEQWSESSISIKAAQAVANKALTESTLPKVPQPKSVNNNNPGSITPVELNG
LAMRGEPAIYPLDPKPPNPYRANFRNGYNORHCVPIKIFYVLTVGNVEFGE
GTRDAAARNNAAMKALQALONPPIPERSPONGESGDDDDDDANKSELSVNEIKL
RMVYEFVYIKESGPHMKSFYRVSVGFSAEGSGSKLSKKRAATVLOEIKLP
PLPYVEKRLFFKKRPKRTIVKAGPEGGGMNPISRALQLOAKKKEPPYVLSEGM
PKRREVMQVGVNEVATGTGNKTIKAKNAEAMLLQIGYASTNLQDLEKTEGK
GMSRPGPEPTNTPKGIILSPDYVDEMASSRPIVSGTGLGLSKDKMNQPSSS
FPSISPTSNSSATIRALLMNGTSSTAFAIIGKSSPPTPCSPVGSKOLEVLIARIQ

```

```

FOAALSALKQFSEGGIDPIDGAMNIEKSLERQAKHLREKADNNOAPGSIADCKKS
NSAV"
BASE COUNT 970 a 605 c 643 g 750 t
ORIGIN
alignment_scores:
Quality: 1067.50 Length: 523
Ratio: 2.885 Gaps: 19
Percent Similarity: 70.746 Percent Identity: 49.331
alignment_block:
us-09-316-048-6 x AK002152 ..
Align seg 1/1 to: AK002152 from: 1 to: 2968
1 MetSerGlnValGlnValGlnValGlnAsnProSerAlaAlaLeuSerG1 17
||| ::::::::::: ::::::::::: ::::::::::: :::::::::::
369 ATGTTCTCAGTGAGCTGAGCTGTTGTTGAGCAGACATGGGAATCCGAAG 418
17 ySerGlnLeuAsnLysAsnGlnSerLeuLeuSerGlnProLeuMetS 34
||| ::::::::::: ::::::::::: ::::::::::: :::::::::::
419 CAGCAGTATAAGAGAGCTCAGCAGAGCTGTTGCCATTAACCTTTG.... 464
34 erLeProSerThrThrSerSerLeuProSerGluAsnAlaGlyArgPro 50
||| ::::::::::: ::::::::::: ::::::::::: :::::::::::
465 .....ACTGAATCTACGGCTTCC.....AAACCA 488
51 lLeGlnSerAlaLeuProSerAlaSerLeuThrSerThrSerAlaAl 67
||| ::::::::::: ::::::::::: ::::::::::: :::::::::::
489 GTTCAAGAGCCA.....CCCAAAAGTAATGTTAAACAATAACCCAGGC.. 530
67 aAlaGlnSerLeuThrProThrValGlnLeuAsnAlaLeuGlyMetLysL 84
||| ::::::::::: ::::::::::: ::::::::::: :::::::::::
531 .....AGTATACTCCAACTGAGAGCTGAGAGCTGCTGCTATGAAGA 573
84 euGlyLysLysProMetLysLysProValAspProTyrSerArgMetGln 100
||| ::::::::::: ::::::::::: ::::::::::: :::::::::::
574 GGGAGAGACCTGCGCATTTACAGGCCATTAGATCCAAAGCCATTCCCAAT 623
101 SerThrTyrAsnTyrAsnMetArgGlyGlyAlaTyrProProAlaGly 117
||| ::::::::::: ::::::::::: ::::::::::: :::::::::::
624 TATAGAGCTAATTACAACTTCGG...GGCATGTACAAATCAGAGGTATCA 670
117 eTyrProPheProValProProLeuLeuTyrGlnValGlnLeuSerValG 134
||| ::::::::::: ::::::::::: ::::::::::: :::::::::::
671 T.....TGCCAGTGCTCTAAGATCTTTAT...GTCAAGCTACCTGTAG 711
134 lGlyGlnGlnPheAsnGlyLysGlyLysThrArgGlnAlaAlaLysHis 150
||| ::::::::::: ::::::::::: ::::::::::: :::::::::::
712 GAATTAATGAATTTTGGGAGAGAAAGACTCGACAGAGCTGTACAC 761
151 AspAlaAlaLysAlaLeuArgLLeuGlnAsnGlnProLeuProG1 167
||| ::::::::::: ::::::::::: ::::::::::: :::::::::::
762 AATGCTGCAATGAAGAGCCCTCCAAAGCTCGAGAAATGAACCTATCCAGA 811
167 uArgLeuAluValAsnGlyArgGlu.....SerGluGluGlu 180
||| ::::::::::: ::::::::::: ::::::::::: :::::::::::
812 AAGATCTCTCAGAAATGTTGAATCAGAAAGCAATATGATGTGCAAG 861
180 snLeuAsnLysSerGluLeuSerGlnValPheGluLeuAlaLeuLysArg 196
||| ::::::::::: ::::::::::: ::::::::::: :::::::::::
862 ATGCAAAATAGCTGAGATCAGCTTACGCTTTGAATATGCTCTGAAGCA 911
912 AATATGCTGTGACGTTTGTAGGTTATTAAGAAAGTGAACACACACATAT 961
213 lLysAsnPheValThrLysValSerValGlyGlnPheValGlyGluGly 230
||| ::::::::::: ::::::::::: ::::::::::: :::::::::::
962 GAAAGCTTTTACTCGAGTCTAGTGAAGAGATTTCTCTCAGAGAGAG 1011
230 lGlyLysSerLysLysLeuSerLysLysAsnAlaAlaAlaValLeu 246

```

[illegible]

```

seq_documentation_block:
LOCUS      AC063967      225016 bp      DNA      HTG      06-SEP-2000
DEFINITION Mus musculus clone RP23-41G23, WORKING DRAFT SEQUENCE, 13 unordered
pieces.
ACCESSION  AC063967
VERSION    AC063967.3 GI:9972323
KEYWORDS   HTG, HTGS_PHASE1, HTGS_DRAFT.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE  1 (bases 1 to 225016)
AUTHORS   McCombie,W.R., Baker,J.P., Bahret,A., Bal,H., Deathia,N.N., de la
            Baatle,M., Huang,E.N., Kling,L., Kirchoff,K.A., Miller,B.,
            Nascimento,L.U., O'Shaughnessy,A.L., Preston,R.R., Rodriguez,M.A.,
            Shah,R.S., Shekher,M., Spiegel,L.A., Toth,K. and VIL,M.D.
TITLE      Mouse genomic sequence
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 225016)
AUTHORS   McCombie,W.R.
TITLE      Direct Submission
JOURNAL    Submitted (22-APR-2000) Lita Annenberg Hazen Genome Sequencing
            Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
            Harbor, NY 11724, USA
            On Sep 6, 2000 this sequence version replaced gi:8439872.
COMMENT    Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
            Laboratory
            Center code: CSHL
            Web site: http://www.cshl.org/genseq
            Contact: mcombie@cshl.org
            ----- Project Information
            Center project name: RP23-41G23
            Center clone name: RP23-41G23
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            1
            121343: contig of 121343 bp in length
            *
            121344 121985: gap of unknown length
            *
            121986 139930: contig of 17945 bp in length
            *
            139931 140572: gap of unknown length
            *
            140573 155472: contig of 14900 bp in length
            *
            155473 156114: gap of unknown length
            *
            156115 167728: contig of 11614 bp in length
            *
            167729 168370: gap of unknown length
            *
            168371 177185: contig of 8815 bp in length
            *
            177186 177827: gap of unknown length
            *
            177828 185945: contig of 8118 bp in length
            *
            185946 186587: gap of unknown length
            *
            186588 194327: contig of 7740 bp in length
            *
            194328 194969: gap of unknown length
            *
            194970 202341: contig of 7372 bp in length
            *
            202342 202983: gap of unknown length
            *
            202984 210036: contig of 7053 bp in length
            *
            210037 210677: gap of unknown length
            *
            210678 216586: contig of 5909 bp in length
            *
            216587 217227: gap of unknown length
            *
            217228 220778: contig of 3551 bp in length
            *
            220779 221419: gap of unknown length
            *
            221420 223928: contig of 2309 bp in length
            *
            223929 224569: gap of unknown length
            *
            224570 225016: contig of 447 bp in length.
            *
            Location/Qualifiers
            1. 225016
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="RP23-41G23"

```

BASE COUNT 59632 a 50195 c 50037 g 57445 t 7707 others  
ORIGIN

alignment\_scores:  
Quality: 921.00 Length: 1792  
Ratio: 2.544 Gaps: 11  
Percent Similarity: 20.201 Percent Identity: 18.694

alignment\_block:  
US-09-316-048-6 x AC063967/rev ..

Align seg 1/1 to reverse of: AC063967 from: 1 to: 225016

```
203 GLUValAlaIArgIuSerGIyProHIsMetIysAsnPhveAlThrly 219
      :::::::::::::::::::::::::::::::::::::::::::::::
89063 CAGGTGCCCCGGAGAGTGGCCCCACACATGAGAACTTTGTGACGAG 89014
      :::::::::::::::::::::::::::::::::::::::::::::::
219 sValSerValGIyGIuPhveValGIyGIuGIyLysSerIyLysI 236
      :::::::::::::::::::::::::::::::::::::::::::::::
89013 GGTTCAGTTGGGGAATTTGTAGGGAGAGGAAGGAAAGCAAGAGA 88964
      :::::::::::::::::::::::::::::::::::::::::::::::
236 IeSerIyLysAsnAlaIAlaIleAlValIeuGIuLysLysLys 252
      :::::::::::::::::::::::::::::::::::::::::::::::
88963 TCTCCAGAAAGAAATGCGCGCAGGCGCTGCTGAGCAGCTTAGAGGCTG 88914
      :::::::::::::::::::::::::::::::::::::::::::::::
253 ProProIeuProAlaValGIuArGIyValIysProArGIleIyLysIyLysTh 269
      :::::::::::::::::::::::::::::::::::::::::::::::
88913 CCACCCCTCCTGCTGTGAGCGAGTGAGCCGGAATCAAGAGAAAG 88864
      :::::::::::::::::::::::::::::::::::::::::::::::
269 rIySPro..... 271
      ::::::::::
88863 TCAGCCCACTGCAGAGGTGAGAGCTTCCCAAGAAACCAACAGCAAG 88814
      :::::::::::::::::::::::::::::::::::::::::::::::
271 ..... 271
88813 AGAGAGCCGGCTGCTGCTGTGGGTAGAGGAGTCTGCTCAGCACC 88764
      :::::::::::::::::::::::::::::::::::::::::::::::
271 ..... 271
88763 ACACGGAGCAAAATGATGTTTACAGACACACATAGAGCTATAGGTTG 88714
      :::::::::::::::::::::::::::::::::::::::::::::::
271 ..... 271
88713 AGGATAATATAGCACCTGATGGGTGAGAGAAAGCTGTCCAGGTGTCA 88664
      :::::::::::::::::::::::::::::::::::::::::::::::
271 ..... 271
88663 GCTTCCCTGGGGTTGGGCACACAGCGCTGTGGTCTTTGTGTGAAGG 88614
      :::::::::::::::::::::::::::::::::::::::::::::::
271 ..... 271
88613 GTAGGGTTTCTTGAGCCAACTTAGCTAGGGATCAAGTGGAGCAGCA 88564
      :::::::::::::::::::::::::::::::::::::::::::::::
271 ..... 271
88563 CTGAGGGCCACAGAGACTATTCCAAAGACTTGATGCTTGCAATGAAC 88514
      :::::::::::::::::::::::::::::::::::::::::::::::
271 ..... 271
88513 TTAGTGGCTTAGTTAATTAATATACAAAGTGGGATGTGGCGCAC 88464
      :::::::::::::::::::::::::::::::::::::::::::::::
271 ..... 271
88463 GCCTTTAATCCACAGCTTGGGCGGAGAGGAGCAACAATTTCTGAGTTC 88414
      :::::::::::::::::::::::::::::::::::::::::::::::
271 ..... 271
88413 GAGGCCAGCTGGTCTACAGAGTAGTTCCAGGACAGCGAGGCTACACA 88364
      :::::::::::::::::::::::::::::::::::::::::::::::
271 ..... 271
88363 GAGAAACCTGTCTCGGGAATAAATAATATACAGATTTCACATGAGTA 88314
```

```
271 ..... 271
88313 CTGAGACTGAGAGGTGGCTCAGCAGTTAAAGCACTGGCTGCTAGC 88264
      :::::::::::::::::::::::::::::::::::::::::::::::
271 ..... 271
88263 AGAGACCTGGGTTCAACAATCTGTGTACACAGATTACAGACTCTCTCT 88214
      :::::::::::::::::::::::::::::::::::::::::::::::
271 ..... 271
88213 GAACCTCCTGAGCACCAGTAGTCACATGTGAATGATGACAGACAAACAC 88164
      :::::::::::::::::::::::::::::::::::::::::::::::
271 ..... 271
88163 TCACATACAGAGATAGGAATCTTTATTACACATGTGGCCCTTTCCGC 88114
      :::::::::::::::::::::::::::::::::::::::::::::::
271 ..... 271
88113 ACTAGAAAGGAGACAGACTTCAGCCAACTCTCCTTTTTCATCTGTGTG 88064
      :::::::::::::::::::::::::::::::::::::::::::::::
271 ..... 271
88063 TCTGAAATTAATCTTCATCATTCCTGACAGTGCCATGGGAAAGAGG 88014
      :::::::::::::::::::::::::::::::::::::::::::::::
272 ..... 275
88013 AGCCCTTAGTCTCCTCAGAAAGCTTGCAATCCTCTCTCCCGGGCCTCAC 87964
      :::::::::::::::::::::::::::::::::::::::::::::::
275 ..... 275
87963 TGGAGGTTTTTAATTCATAACAGTGTGTTGTTTGTCTACTGGGATTG 87914
      :::::::::::::::::::::::::::::::::::::::::::::::
275 ..... 275
87913 GATGAGAGACTCAGGTGTGTGACCAAGCAAGCCTTCATAGGATG 87864
      :::::::::::::::::::::::::::::::::::::::::::::::
275 ..... 275
87863 CATCTTCACTGAGGCTTACAAAGCTTTAAGTGTCTGTGGTTT 87814
      :::::::::::::::::::::::::::::::::::::::::::::::
275 ..... 275
87813 AATGATTGAGAAAGCTAGGCATGGGTGCCATGTAGATTCAATGACGCA 87764
      :::::::::::::::::::::::::::::::::::::::::::::::
275 ..... 275
87763 GGTCAAGCCCTGGTTCATCCTCCCTTTTGTGATCTTTTTTTTTT 87714
      :::::::::::::::::::::::::::::::::::::::::::::::
275 ..... 275
87713 TTTTCTGAGACTGCTGTGTAGCAGGCTGGCTCGAATTCAGAAAT 87664
      :::::::::::::::::::::::::::::::::::::::::::::::
275 ..... 275
87663 CCGCTGCTCCACCTCCCAAGTCTGSGATTAAAGSGGTGACACACAC 87614
      :::::::::::::::::::::::::::::::::::::::::::::::
275 ..... 275
87613 TGCCGGCTCTTTTGTGATCTGTGCGCCAGTGTGCTTACAGCTTTT 87564
      :::::::::::::::::::::::::::::::::::::::::::::::
275 ..... 275
87563 ACATGGCAGTTCACCTGTGTAATGTCCCAAGTATGATGAGTACGCTTC 87514
      :::::::::::::::::::::::::::::::::::::::::::::::
275 ..... 275
87513 CGCATGCCGAGAGTACAGAGTCAACAGCAACACCAACCAAGAGCA 87464
      :::::::::::::::::::::::::::::::::::::::::::::::
275 ..... 275
87463 TCTGAGCATGATCCTGAACATGTTTGTAGTGTAGAGAGGTTGTTCT 87414
```

```
275 ..... 275
87413 GAGCGTTTACACGACGACTAGAGTTTCTGTCTAGTCCCTGCTGAC 87364
275 ..... 275
87363 GCTGACGTGCTGTCTCTCTGTGAGTAACAGCATATTGTCTCTGCAC 87314
276 ..... GlnThrSerProGluTyrGlyInglYIleasn 286
87313 TCGTTTCTTTTTCAGCTACAGACAGCCCCGATTATGGCCAGGGATGAAT 87264
287 ProIleSerArgLeuAlaInllegInlAlaLysLysLysGluYr 303
87263 CCTATTAGTAGACTTGCACAGATCCAGACGCAAAAAAGAGAGAGAGCC 87214
303 oGluTyrThrLeuLeuThrGluArgGlyLeuProArgArgArgGluPheY 320
87213 AGAGTACATGCTCTCTTACAGAACGAGGTCTTCCAGTCGCGAGGAGTTTG 87164
320 alMet..... 321
87163 TGAATGCAAGTGAAGTTGGGGTTGAGCAGCCCTTCCCTGACCATGTC 87114
321 ..... 321
87113 CTAAATCGGGAGTGTTCAGATTGAGTGAAGTTGAGTGCCTGCTTG 87064
321 ..... 321
87063 CTCTCTGAGGCCAAGCAAGCGCAGGTGGATCTATGAGTTCAAGAC 87014
321 ..... 321
87013 TCCACATTAGTAGATCCCATCAGAAACAGATCCACTCTGTAAAGAT 86964
321 ..... 321
86963 GAGTTGGACACAGATTGGCTGAAGTGGGCGAGAGCTTTGATGAAG 86914
321 ..... 321
86913 TGTGTGAATTGTTCACAGACGCCCTCAGAGAGTGTATGAGAG 86864
321 ..... 321
86863 AGAGCCCCACTGCATCTAGATATATACCTGTCTAATAATTGATGAT 86814
321 ..... 321
86813 GTCGTGGCACACATTTTCATACATAGTTTTTAAGTTTAAAGAAC 86764
321 ..... 321
86763 TCTTTAAGACGTGTGAGATGGCTCAGTGGTTAAGACACTGACTGCTCT 86714
321 ..... 321
86713 TCCGAAGTCTGAGTTCAATCCACGACCAATGTGTGCTCACAACC 86664
321 ..... 321
86663 ACCCATATGAGATCTGATGCCCTTCTGTGTCATCTGAAGACAGCTAG 86614
86613 AGTGTACTTAATAAAGTCTTAAAAAAATAAAGTCTTAAAAATAGA 86564
321 ..... 321
86563 GTTTTCATATCTCTAACGTGTGTGGCCATCAGCAGTAGTAGTTGA 86514
321 ..... 321

86513 GAATACGAAACATAAAGCTTTTACCATCAAGTTCAGACATCATCCCTGG 86464
322 ..... GlnValLysY 325
86463 GGCTTTGGGTTAAGTACAGAGTCTTTGTTGGTGTTTTACACAGATGAAG 86414
325 alGlyAsnHisThrAlaGluGlyThrGlyThrAsnLysLysValAlaLys 341
86413 TTGGCATCAGACTGCCAAGAGAGTGGTACCAATAGAAGGTGGCCAG 86364
342 ArgAsnAlaAlaGluAsnMetLeuGluIleLeuGlyPheLysValProG1 358
86363 CATAATGCTGTGAGAAACATGCTGAGATCTGTGGGTTCAAAAGTCCCA 86314
358 nArgGlnProThrLysProAlaLeuLysSerGluLys..... 371
86313 GCGCAGCCTGCCAACCCAGCACTCAAAATCAGAAAGAGAGTACTGTG 86264
371 ..... 371
86263 GGACTCTGTGCCAGTGGGCTTAACCTCCATGCGGCAAGCCTGGGGCAA 86214
371 ..... 371
86213 GGCTCATCTCCCTCAGTAAGCGGCTGCCCCCGGGGCTCTCCATT 86164
371 ..... 371
86163 GAGATTAGATTGACAGAGGAGATTCTGTCCAGCTGCATCATCTGG 86114
371 ..... 371
86113 GAGCCATCCATTGCTTAAGTCTCCTGTTAGTGTACTATAGGGCTAGAT 86064
371 ..... 371
86063 GTCATCTCCGAGAGAGTCTCTCTCTACTTGTATGAGTGGAGTGGG 86014
371 ..... 371
86013 TTGAGAGGAATAAATGACTGCCACCTGTCTGCATGACCCAGTTTGA 85964
371 ..... 371
85963 TATGCTGTGTTTCTGTGGACAAAGCTAGCATGTCAATAAGCCAAAT 85914
371 ..... 371
85913 CTATGATTTGAGTGAATATCATATACGCCACCATTTGCTGCTACAG 85864
371 ..... 371
85863 AGTTGTCATATTGGGATTGAGCTTCTGAGCTCTGCCAGTTTCATTACA 85814
371 ..... 371
85813 AGAATAGTCCCTCCACTGCTGGAGCTTACTGTGTACTGCTGCCCTT 85764
371 ..... 371
85763 GATTGGCTTCAGATTTCATTCCAAAACAGTGTAGAGTCTATTAGTT 85714
371 ..... 371
85713 CAAGACTCACATAGTAGATCCGTCACAGAAACAGTATCCGCTCTGTA 85664
371 ..... 371
85663 AGAGTAGTTGGACACAGAGTTTGCTGAAGTGGCTGGGGACAGCTTTGA 85614
371 ..... 371
```

85613 TGAAGTGTGCTGGAATCTGTCTTCTTCTGTATAGCATTAGGTACCTC 85564  
371 ..... 371  
85563 TGTGTCTTGTGGATGTACTGTTTGGATGTTCTAGTTATATGAAGTG 85514  
371 ..... 371  
85513 TTCTCTTAGCACAGTGCAGACCTACTTCCCAAGAAGCTGAACACT 85464  
371 ..... 371  
85463 TTTAAGCCAGACATTTCTCTGCGAGTGTTCGTGTTCTTGACACAGT 85414  
371 ..... 371  
85413 GCCAGTATGCAATTCAGATCTGCTGTGCAATGCTGCCATGGGAGG 85364  
371 ..... 371  
85363 CGGAGGCTTGTCTATCAACAGACCCAGTGTGTATGCTCACT 85314  
372 ..... ThrProIleLysLysProGlyAspGlyArgL 382  
85313 GCCATTTTCTTCTTAAAGACTCCAGTAAAGAAACAGAGACGGAAGGA 85264  
382 ySValThrPhePheGluProGlySerGlyAspGluAsnGlyThr..... 396  
85263 AAGTAACGTTTTTGAACCTAGCCCTGGGGATGAAATGGAATAGTAAG 85214  
396 ..... 396  
85213 TGTGGCCCTCTGTCAATGACCAATGTGCTGTGCTTGTGTGCTCT 85164  
396 ..... 396  
85163 CTGTAGTTTTTGAAGCAGAGTATTCCTGCTCTGTATGCTGGAT 85114  
396 ..... 396  
85113 GGTAGGGTGGCTGTAGCTTCTGGGTGGTATGCTGACTTGGTGAC 85064  
397 ..... Ser..AsnLysGluAspGluPheArgMetProTyrLeu 408  
85063 CTGTGTCCCTCGCAGGTAAAGAGCAGAGGAGTGCAGATGCCATTACTT 85014  
409 SerHisGlnGlnLeuProAlaGlyIleLeuProMetValProGluValAl 425  
85013 AGCCATCAGCAGCTGCCAGCTGGAATTCGCCCATGTGCGCGAAGTTGC 84964  
425 agLhlaValGlyValSerGlnGlyHisHisThrLysAspPheThrArg 442  
84963 CCAGGCTGTGGGGTGTATGTCAGAGCACACACCAAGATTTCACACAGG 84914  
442 laaIaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArgGlu 458  
84913 CAGCTCCAAATCCTCGCAAGCAGCAGGTAATGTCATGATAGCCCGAAG 84864  
459 LeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysAsnAs 475  
84863 TTGTGTGACGGGGGACCTCCGCCACAGCCAGACCATTTTAAAGAGTAA 84814  
475 nIleSerSerGlyHisValProHisGlyProLeuThrArgProSerGluG 492  
84813 CATCTCTTAGCCAGCCATGCCATGACCTCGCAGCATAGACCTCTGAGC 84764  
492 InLeuAspTyrLeuSerArgValGlnGlyPhe..Gln..... 503  
84763 AACTGTACTACTTTCAGAGCCAGGATTCAGTAAACGTTGTTGTCGA 84714  
503 ..... 503  
84713 TCCTGTGGCCTTACGTTCTTCTGTGACAGAAATGCTGTGTAGTGTGTC 84664

503 ..... 503  
84663 AAGACCCGTGGTCTGTCTCAGCAGCTACAGAAACAGAAAAGCAAGAT 84614  
503 ..... 503  
84613 TATATGCTTATTTAAGATTTGAATGTACAGTGAAGTTCTGAAGAAAGA 84564  
503 ..... 503  
84563 AGATGTGGATGAGAAATGCTTATATAAGTAAAGTATTCTTAAGCCC 84514  
503 ..... 503  
84513 AACACCTTGTAGTATGATTACTGGTGAAGAACAACCACTCCAGCAAGTTG 84464  
503 ..... 503  
84463 TCCTCAGTGTGCACACCTCGCCAGAGGTTTACGTGTGCGCCAGGGCTG 84414  
503 ..... 503  
84413 AAGGAGAGAAAGACACTTCTTACTTAGAAGCAGCCACTACATGAGCATGG 84364  
503 ..... 503  
84363 GAGAGAGGCTTGAGGCTTCTGGGCTGGGGCCACATGCAAGAGTTGCT 84314  
504 ..... ValGluTyrLysAspPheProLysAsnAsnLysAsnG 516  
84313 GTTTTTCCTTAGTGAATACAAAGATTTTCCCAAGAACACACAGAAACGA 84264  
516 uPheValSerIleuIleAsnGlySerSerGlnProProLeuIleSerHisG 533  
84263 GTGTATCTCTCATCAACTGCTCTCCACAGCCGCTCTGTCAGTCAATG 84214  
533 LylIleGlyLysAspValGluSerCysHisAspMet..... 544  
84213 GCATCGCAAGATGTGGAGTCTGTCTCATGATATGTATGTGGGCCAGG 84164  
544 ..... 544  
84163 AGGTGTGAGAAGAGCTGCGACATTCCTGTTCAGTCAGTGTAGTG 84114  
544 ..... 544  
84113 GCCTCAGAAAGAGGACATGGTTCCTCTAGCCTTGTGACATGTCTGA 84064  
544 ..... 544  
84063 TTTGATTCAATTATTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 84014  
544 ..... 544  
84013 GGCAAGCTTAGAAAGCCCCCTTTGGGAGATTCTGTCAAAGAAGAAATCA 83964  
544 ..... 544  
83963 GCACACTTGAGCTTGTGTTCACCTCGATGGAGCCATATGCAACCCCTTA 83914  
544 ..... 544  
83913 AGCGGCAACCTCTGATTTTGAAGGAAACATGGAGTTTGATGTTGAACA 83864  
544 ..... 544  
83863 GGATTCAAACATTTCAATGGAAGTTTAAAGAGCTGAGTCTGTTTC 83814  
545 ..... Ala..AlaLeuAsnI 549  
83813 CCTGTGTTGATTATTCCTTCTTCACTTTCAGCTTTCAGGCTGACGTGAACAT 83764

549 e1eulysleu1eusergl1eulasp1ng1nserthr1glumeteproragt 566  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 83763 TTTAAAGCTCTGTCGATGTGGACCAACAGACAGACATGCGAAGAA 83714  
 566 hrc1yasn1ng1yPrometSer 572  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 83713 CAGGAATGACACAGCTTCA 83694

seq\_name: gb\_prf:HSBB140E4

seq\_documentation\_block:

LOCUS HSBB140E4 163253 bp DNA PRI 07-FEB-2000  
 DEFINITION Human DNA sequence from clone RP13-140E4 on chromosome X, complete  
 sequence.

ACCESSION AL121877

VERSION AL121877.13 GI:6933862

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 163253)

AUTHORS Heath,P.

TITLE Direct Submission

JOURNAL Submitted (07-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

#### COMMENT

On Feb 7, 2000 this sequence version replaced gi:6911935.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.

This sequence has been finished according to sequence map criteria  
 as follows. An attempt is made to resolve all sequencing problems,  
 such as compressions and repeats, but not necessarily within known  
 annotated human repeat sequence elements (e.g. Alu). Where the  
 sequence is ambiguous, there is an annotation using the 'unsure'  
 feature key.

The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information  
 on the WORMPEP database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence

was generated from part of bacterial clone configs of human  
 chromosome X, constructed by the Sanger Centre Chromosome X Mapping  
 Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/ChrX>

RP13-140E4 is from the library RPCR-13.1 constructed at the Roswell  
 Park Cancer Institute by the group of Pieter de Jong. For further  
 details see <http://bacpac.med.buffalo.edu/>  
 VECTOR: PACe3.6

This sequence is the entire insert of clone RP13-140E4.

#### FEATURES

##### source

1. 163253  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="X"  
 /clone="RP13-140E4"  
 /clone\_1b="PCR-13.1"  
 25. 188  
 /note="L1M4 repeat: matches 5114. .5296 of consensus"  
 repeat\_region  
 184. 411  
 /note="L1P3 repeat: matches 2841. .3072 of consensus"  
 repeat\_region  
 412. 694  
 /note="AluSq repeat: matches 1. .287 of consensus"  
 repeat\_region  
 695. 3450  
 /note="L1P3 repeat: matches 232. .2841 of consensus"  
 repeat\_region  
 3451. 6427  
 /note="L1PA2 repeat: matches 3148. .6144 of consensus"  
 repeat\_region  
 6432. 7143  
 /note="L1MD1 repeat: matches 5283. .6016 of consensus"  
 repeat\_region

repeat\_region  
 7164. 7316  
 /note="Aluub repeat: matches 137. .290 of consensus"  
 repeat\_region  
 7339. 7646  
 /note="L1MD2 repeat: matches 6016. .6330 of consensus"  
 repeat\_region  
 7650. 7754  
 /note="L1MC5 repeat: matches 7608. .7714 of consensus"  
 repeat\_region  
 7767. 8072  
 /note="AluY repeat: matches 1. .305 of consensus"  
 repeat\_region  
 8095. 8136  
 /note="21 copies 2 mer ca 90% conserved"  
 repeat\_region  
 8152. 8245  
 /note="L1MC5 repeat: matches 7478. .7572 of consensus"  
 repeat\_region  
 8273. 8494  
 /note="AluYo repeat: matches 84. .305 of consensus"  
 repeat\_region  
 8503. 8633  
 /note="L1MC5 repeat: matches 7305. .7446 of consensus"  
 repeat\_region  
 8712. 8879  
 /note="84 copies 2 mer tt 78% conserved"  
 repeat\_region  
 8899. 9190  
 /note="AluYo repeat: matches 8. .300 of consensus"  
 repeat\_region  
 complement(16833. .17221)  
 /note="match: STS: Em:LA1960; match: STS: Em:LA1960"  
 repeat\_region  
 12531. 13017  
 /note="L2 repeat: matches 2083. .2505 of consensus"  
 repeat\_region  
 13018. 13320  
 /note="L1M1 repeat: matches 6004. .6303 of consensus"  
 repeat\_region  
 13321. 14004  
 /note="L2 repeat: matches 1197. .2083 of consensus"  
 repeat\_region  
 complement(16833. .17221)  
 /note="match: GSS: Em:AQ558019"  
 repeat\_region  
 16840. 17277  
 /note="match: GSS: Em:B65742"  
 repeat\_region  
 complement(16845. .17241)  
 /note="match: GSS: Em:AQ149727"  
 repeat\_region  
 complement(16848. .17220)  
 /note="match: GSS: Em:AQ240745 Em:AQ474471"  
 repeat\_region  
 16886. 17629  
 /note="MER4D repeat: matches 234. .973 of consensus"  
 repeat\_region  
 18183. 18407  
 /note="L1P3 repeat: matches 5928. .6148 of consensus"  
 repeat\_region  
 18455. 19002  
 /note="MER41A repeat: matches 1. .554 of consensus"  
 repeat\_region  
 19289. 19465  
 /note="MIR repeat: matches 15. .198 of consensus"  
 repeat\_region  
 19583. 20000  
 /note="match: GSS: Em:AQ017861"  
 repeat\_region  
 19875. 20274  
 /note="MER57B repeat: matches 1. .403 of consensus"  
 repeat\_region  
 20275. 21084  
 /note="MER57-Internal repeat: matches 1. .806 of consensus"  
 repeat\_region  
 21078. 22422  
 /note="MER57-Internal repeat: matches 440. .2799 of  
 consensus"  
 repeat\_region  
 22425. 22764  
 /note="MER57-Internal repeat: matches 3892. .4230 of  
 consensus"  
 repeat\_region  
 22765. 23041  
 /note="AluX repeat: matches 8. .286 of consensus"  
 repeat\_region  
 23042. 24919  
 /note="MER57-Internal repeat: matches 4230. .6114 of  
 consensus"  
 repeat\_region  
 24939. 25118  
 /note="L1P1 repeat: matches 5980. .6153 of consensus"  
 repeat\_region  
 25119. 25422  
 /note="AluY repeat: matches 1. .304 of consensus"  
 repeat\_region  
 25423. 25607  
 /note="L1P1 repeat: matches 5777. .5980 of consensus"  
 repeat\_region  
 25635. 25672  
 /note="19 copies 2 mer ta 78% conserved"  
 repeat\_region  
 25681. 26070  
 /note="L1P1 repeat: matches 5397. .5790 of consensus"  
 repeat\_region  
 26082. 27136  
 /note="MER57-Internal repeat: matches 6489. .7537 of  
 consensus"  
 repeat\_region



```

159673 GAATA. GGGCCAGGATGAAACCTATTAGCTGCCGGCCAAATTCACA 159721
296 nAlaLysLysGluLysGluProGluThrLeuLeuThrGluArgGlyL 313
||||| ||||||| ||||||| ||||||| ||||||| |||||||
159722 GGGCAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 159771
313 eubPrgArGArGArGValMetGlnValLysValGlyAsnHisThr 329
||||| ||||||| ||||||| ||||||| ||||||| |||||||
159772 TGCGTCGATGTCGAGAAATTTGTGATGATGAAGTAGCAATGAAGCT 159821
330 AlaGluGlyThrGlyThrAsnLysVal...Ala.LysArgAsnAla 345
||||| ||||||| ||||||| ||||||| ||||||| |||||||
159822 GCTACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 159871
345 IaGluAsnMetLeuGluLeuGlyPheLysValProGlnArgGlnPro 361
||||| ||||||| ||||||| ||||||| ||||||| |||||||
159872 CAGTACGATGCTATTACAACTGGTTATTAAGCA..... 159906
362 ThrLysProAlaLeuLysSerGluGluLysThrProLysLysProG 378
||||| ||||||| ||||||| ||||||| ||||||| |||||||
159907 .....TCCATTATCTTCAGATCA.....CTTGAGAAAGACAG 159941
378 yAspGlyArgLysValThrPhePheGluProGlySerGlyAspGlu 395
||||| ||||||| ||||||| ||||||| ||||||| |||||||
159942 GGAAGAACAA.....GGATGGAGTGTCCACAGCCCTGGCTTCC 159985
395 LysSerAsnLysGluAspGluPheArgMetProTyrLeuSerHisGln 411
||||| ||||||| ||||||| ||||||| ||||||| |||||||
159986 CAACAATAT..... 159996
412 GlnLeuProAlaGlyLeuLeuProMetValProGluValAlaGlnAl 428
||||| ||||||| ||||||| ||||||| ||||||| |||||||
159997 ...ACGCCAAAGAGATTTCTTCATTTGTCTCCGATGTTTATCA 160043
428 LglValSerGlnGlyHisHisThrLysAspPheThrArgAlaAla 445
||||| ||||||| ||||||| ||||||| ||||||| |||||||
160044 GGAAGCCAGCCACACAAAGTAACTGTGCTACTCTAGCTATTGTT 160093
445 snPro.....Ala 447
||||| ||||||| ||||||| ||||||| ||||||| |||||||
160094 CACCAGAGATTAACCAACCTTCAAGCTTTGTTCAATGATGCC 160143
448 LysAlaThrValThrAlaMetIleAlaArgGluLeuLeuLysGly 464
||||| ||||||| ||||||| ||||||| ||||||| |||||||
160144 ACATCAAAATGTTCACTCAATGCGCAGGAACTCTTATGAATGAA 160193
464 rSerProThrAlaGlu...ThrIleLeuLysAsnAsnIleSerSer 480
||||| ||||||| ||||||| ||||||| ||||||| |||||||
160194 ATCTTCTACAGCTGAAGCCACAGGTTTAAAGCA.....AGTTCT 160237
480 lValProHisGlyProLeuThrArgProSerGluGlnLeuAspTyr 496
||||| ||||||| ||||||| ||||||| ||||||| |||||||
160238 CTCCTCCCTGTTCTCCAGTA...CAACCTTCANAACAACGGAAT 160284
497 SerArgValGlnGlyPheGlnValGluTyrLysAspPheProLys 513
||||| ||||||| ||||||| ||||||| ||||||| |||||||
160285 GCAAGGATTCAGAGCTTTCAGATGATGAATTAAGC.....AAA 160328
513 nLys 514
||||| ||||||| ||||||| ||||||| ||||||| |||||||
160329 AAAA 160332
seq_name: gb_pr4:AL133174
seq_documentation_block:
LOCUS AL133174 113836 bp DNA PRI 14-JUN-2000
DEFINITION Human DNA sequence from clone RP3-470L14 on chromosome 20, complete
sequence.
ACCESSION AL133174
VERSION AL133174.15 GI:8573761
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 113836)
Ramsay, H.
Direct Submission
Submitted (14-JUN-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jun 20, 2000 this sequence version replaced gi:8388434.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WormPEP; Information
on the WormPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
RP3-470L14 is from the library RCI-3 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://dacpac.med.buffalo.edu/
VECTOR: pCYPAC2
IMPORTANT: This sequence is not the entire insert of clone
RP3-470L14 it may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true right end of clone RP3-470L14 is at 113836 in this
sequence. The true right end of clone RP1-1556 is at 100 in this
sequence.
FEATURES
Source
1..113836
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RP3-470L14"
/clone_lib="RCI-3"
3061..3077
/note="Weak data"
95583..95627
/note="Single clone region. pUC subcloned to give 12x
coverage. Assembly confirmed by restriction digest data."
BASE COUNT 33974 a 25115 c 24233 g 30514 t
ORIGIN
alignment_scores:
Quality: 753.00 Length: 429
Ratio: 4.328 Gaps: 3
Percent Similarity: 40.559 Percent Identity: 40.326
alignment_block:
US-09-316-048-6 x AL133174/rev ..
Align seg 1/1 to reverse of: AL133174 from: 1 to: 113836
371 LysThrProLysLysProGlyAspGlyArgLysValThrPhePheG 387
||||| ||||||| ||||||| ||||||| ||||||| |||||||
38483 AAGACACCCATTAAGAAACAGGAGATGAGAAAGTAACCTTTTGA 38434
||||| ||||||| ||||||| ||||||| ||||||| |||||||
387 uProGlySerGlyAspGluAsnGlyThr..... 396
||||| ||||||| ||||||| ||||||| ||||||| |||||||
38433 ACCTGCTCTGGGAGTGAATGGACTAGTAGTGTGATCTTAATATCG 38384

```



```

396 ..... 396
38383 TCTTCTCTGAAGTGATTCGTGGATTTCATGGGCTTGATTAACGCA 38334
396 ..... 396
38333 GGAGCCAGCCAGTGAAGTCCGATGCAAAAGTCTTTATCATTCCTG 38284
396 ..... 396
38283 ATTCCCTGCGCTCCATGCTGCCAAGGGAGGCGCGGTGGGTTGA 38234
396 ..... 396
38233 TGGTGTCATTGCGTTTTTCCAGGCTGTGATGTGATTGCTGTTACCGT 38184
397 ..... Ser..AsnLysGluAspLuphearGmetProTyrLeu 408
38183 GTATGTCCTCCAGGTAATTAAGAGATGATTCAGATGCGCTTATCTA 38134
409 SerHisGlnInleuProAlaGlyLeuProMetValProGluValAl 425
38133 AGTATCAGAGCTGCGCTGGATTCTTCCATGAGTGGCCGAGGTCGC 38084
425 aglAlaValAlGlyValSerGlnHisHisThrLysAspPheThrArg 442
38083 CCAAGGCTGTAGAGTTAGTCAGAGACATCACACCAAGATTTTACAGGG 38034
442 laAlaProAsnProAlaLysAlaThrValThrAlaMetlleAlaArgLu 458
38033 CAGCTCCGAATCTGCGCAAGGCCAGCGTAACCTGATGATGAGCCGAGAG 37984
459 leuLeuTyrGlyGlyThrSerProThrAlaGluThrIleuLysAsnAs 475
37983 TTGTGTATGGGGACCTGCGCCAGACCGAGACCAATTTAAAGATTA 37934
475 nileSerSerGlyHisValProHisGlyProLeuThrArgProSerGlu 492
37933 CATCTCTTCAGGCCAGCTACCCCATGAGACCTCTCAGAGACCCCTGAGC 37884
492 leuLeuAspTyrLeuSerArgValGlnGlyPheGln..... 503
37883 AACTGACATCTTTCCAGAGTCCAGGATTCAGTAACGTCTGCGCT 37834
503 ..... 503
37833 GAGCTGTGATGGCTGTGCTCAGACAGTCATCTTTGGCTTCTCAGA 37784
503 ..... 503
37783 AAATGAACCTTGAACCTTGTACTTTTCTTAATCAAGTCTCTCCAGG 37734
503 ..... 503
37733 GATGGCTGGGAGACAGTATGAAAGGTAGAGGATATAGAATATTGTA 37684
503 ..... 503
37683 AGCATATGTGTAGAGAAATATCTTTATCTTTGGTATGAATTTCAA 37634
503 ..... 503
37633 CATGCAAGCCACCTAAAGCTTCTAACCAGTACAGAGTGGTTAAATT 37584
503 ..... 503
37583 TGGATTCAAGAAGACAGATCATGCTGTGACACTTCTGTCTGTAGGA 37534
503 ..... 503
37533 AAGATATGCGCTTCTCCAGTCTTTGGGCTTCTGTAGACAGACACTAGGTTTC 37484

```

```

503 ..... 503
37483 CTGGAAGTCTTGCGGTGACAGGGCAGGTGTATGATGCCACAGTGGG 37434
503 ..... 503
37433 CTGAGTTGAAAAAAGAGCTGTGATCTTGTGTTGGGCTGAGCTGC 37384
503 ..... 503
37383 AGCCCTCAGCCCTCATCTGCTGCTCCCTACCTAGCTAGCTCTC 37334
504 ..... ValGluTyrLysAspPheProLysAsnAsnLysAsnG 516
37333 TTTTCCCCCTTAGTTGAATACAAAGACTTCCCAAAACAAACAAGAG 37284
516 lupheValSerLeuIleAsnLysSerSerGlnProProLeuIleSerHis 532
37283 AATTGTATCTCTTATCAATTCCTCTCAGCCACCTGTGATGACCAT 37234
533 GlyIleGlyLysAspValGluSerCysHisAspMet 544
37233 GGTATCGCAAGAGTGTGAGCTCTGCCATGATATG 37198

seq_name: gb_in3:DROSTAUEN

seq_documentation_block:
LOCUS DROSTAUEN 5237 bp mRNA INV 26-APR-1993
DEFINITION D.melanogaster mRNA, complete cds.
ACCESSION M69111
VERSION M69111.1 GI:158505
KEYWORDS
SOURCE
ORGANISM Drosophila melanogaster 0-4 hour embryo cDNA to mRNA.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 5237)
AUTHORS Johnston,D.S., Beuchle,D. and Nusslein-Volhard,C.
TITLE Staufen, a gene required to localize maternal RNAs in the
Drosophila egg
JOURNAL Cell 66, 51-53 (1991)
MEDLINE 91300552
FEATURES
source
Location/Qualifiers
1..5237
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/dev_stage="0-4 hour embryo"
<1..5237
275..3355
/codon_start=1
/protein_id="AAA73062.1"
/db_xref="GI:158506"
/translation="MOHNVAARPAPEHRAHHSHSHAMHLHPGMEQHLPSLQOO
OQPPPPQOPRRHDLARLHHNHLLHOQOOOQOOTSNOAAVAALAAAGAAHHGNINSN
GNSISNSNOMOKIROOHOLSSNGLLGNOPGPPPOAFNPPEFAGHLLTAGPGMPQ
PPHMAHNGSYVAAPRPYHYSOAKPAKYNHGSNNSNNSNNTAPYALLONT
YNOKVYVPVQVETPRPYPTTNATNTSNSTVYASBPVOEDTSOKPEPQOE
PASADHVSTGINDATGALSNEDTSSGKGGKDKTPWCVLNELARTNKITHOTRLEE
RPAHCKTFTVLMGDEEVSADGEFKRAQHLAAKALEETMYKHPKIRSEEGG
PMKTHITPTVELNALAMKLGORTFYLDPDQIIPNDISIVPPEFAGHLLTAGPGMPQ
PPPPAYALRORLNGFVPIPSOPMHPEFHGQPPPPKPPFPFAGHLLTAGPGMPQ
GNGPPEPVPTPSPKTTLEFGOKPFGVIGRTLOQAKHDAARLQVLTQOIASPEEA
LSDSMDGSKSPDISOVHEIGIKRMNTYHFKYLREGRPAHMKNETACTIGSIVPEGE
GNGKYSKRAEKRLVELQKLRPLPTPTQPLKRIKYVTPKSGAARAEGVYSTG
GTMQGTCKPERRRLNPPKDKLIDMDADNPITKLIQDQTRKEKEPTEFLAKNGET
ARRREVEVMSASGATAGCTGNSKLAERNAQALELELAQVPTNPNTOSSEECT
SATMSAVTAPAVEATGCKVPMVATPGMPGLILIRQKKPAKKRDOIVIKSNVES
KEEENKAEVAAAEENSNSANSNGSDSOATFAASESALNTSGNSGVSS
NNSNNGANTDGNNAHESKNNTSSNSTNTSAGVHMEKQILYLSKILDPVPSDY
PKGNHNEFLITVLTSHRPOICHGVGKSEESONDAASNALITLSKLGNNMK"
BASE COUNT 1596 a 1291 c 1204 g 1146 t
ORIGIN
Chromosome 55A, right arm.

```

## alignment\_scores:

Quality: 716.00 Length: 644  
Ratio: 2.081 Gaps: 18  
Percent Similarity: 53.416 Percent Identity: 32.143

## alignment\_block:

US-09-316-048-6 x DROSTAUFEIN ..

Align seg 1/1 to: DROSTAUFEIN from: 1 to: 5237

```
71 1lethProthrvAluLeuansAlaLeuCyMeLysLeuGlyLys 87
|||||
1466 ATTACGCCGCGGTGAATTGATGATCCCTGGCCATGAAGTGGAGACAGG 1515
87 sProMeTYr.....LysProValAsp..... 94
:
1516 CACCTTCTACCTGCTGTGATCCACCTCAGATTCTCCACGAGATTCATAG 1565
94 ..... 94
1566 TGCCTCCGCGAGTTGCCGGCGGTCATTACTACTACCTCACCAGACAGG 1615
95 ...ProTYSerArgMetGlnSerThrTYrAsnTYrAsnMetArgLysL 110
|||
1616 ATGCCCGCAACACCTCCACACCTGCTTACGCTCTCGCGCAGCGGCTTG 1665
110 Y..... 110
1666 CAATGATTTTGCCCATTCATCGCAGCGCATCCGATTCCTTCTTC 1715
111 .....AlaTYrProArgTYrPheTYrProPhePro 121
|||||
1716 ATGGACCTGTGACGACCGCTTCTCCGAAATTTCATCAGCCTTTGCG 1765
122 ValProLeuLeu..... 126
|||
1766 CTGCCACCGCCTCTCGAGCTCATGTACATCATGACCGAATGGCCATT 1815
127 .....TYrGlnValGlnLeuSerValGlyGly 136
1816 TCCATCCGTGCTACGCCACCCAGCAAAATCACCCCTGTTGTGGCAAGC 1865
136 1ng1nPheasnGlyLysGlyLysThrArgGlnAlaAlaLysHisAspAla 152
|||
1866 AGAAGTTTGTGGCATCGACGACACACTGCAGAACGAAACATGACGCT 1915
153 AlaAlaLysAlaLeuArgGlyLeuGlnAsnGluProLeu.....Pr 166
|||||
1916 GCGCGCAGAGCCTGCAGGTTCTCAAGACACAGCCCATATCTGCCTCTGA 1965
166 og1uAtgLeuGluValAsnGlyArgGluSerGluGluGluAsnLeu 183
|||
1966 GAGGCTCTCGAG.....GATTCGATGATGAGGGTGACAAAGA 2003
183 ySerGluLeuSerGlnValPheGluAlaLeuLysArgAsnLeuPro 199
|||||
2004 AGTCGCCCATCTCGCAGGTTCAAGAGATCGAATTAAGCGCAACATGAC 2053
200 ValAsnPheGluValAlaArgGluSerGlyProProHisMetLysAsn 216
|||||
2054 GTTCATTTTAAGGTGTGGCGAGAGGCGCGGCACATGAAAGAACTT 2103
216 eValThrLysValSerValGlyLupheValGlyGluGlyLys 233
|||||
2104 TATACAGCCTCATGTTGTGGATCTATTCTACCGCAAGAGAAAGAAAG 2153
233 eTYrLysLysSerLysAsnAlaAlaLeuLysGluGluLeu 249
|||||
2154 GCAAAAAGATGTCAGAAAGCGTGGCGGAAAGATGCTCTGCTGAGTTG 2203
250 LysLysLeuProLeu...ProAlaValGluArgValLysProArgL 265
```

```
2204 CAAAGCTGCGCCCTTACGCCACCAAGCAGACGCAATTGAAGCGAT 2253
|||
265 eLysLysLysThr..... 269
2254 TAAAGTTAAGACACACGAGCAAAAGTGTCAGCTGCCAGAAAGTTCCG 2303
270 .....LysProLeuValLys 274
2304 TTGTGCGGGCACTGATGTACTATGCAACGGGAAAGCCGAGAGACGT 2353
275 ProGlnThrSerPro.....GluTYrGlyGlnGly 285
|||
2354 AAAGCGCTCAATCCGCCGAGAGCAAGCTGATAGACATGAGACATCGGA 2403
285 eAsnProLysSerArgLeuAlaGlnLeuGlnAlaLysLysGlyLys 302
|||||
2404 TATCCCATTTACAAAGCTTAATTCACCTGACGACAGACGCGTGAAGAAAG 2453
302 1uProGluTYrThrLeuLeuThrGluArgGly.....LeuProArgArg 316
|||
2454 AACCCATTTTCAGCTGATAGCCAAAACGGCAATAAACCCCTGACGA 2503
317 ArgGluPheValMetGlnValLysValGlyAsnHisThrAlaGluLys 333
|||||
2504 CGGGAATTCGTTATGAGAGTCTCCGCCAGCGAGACAGACCCGTGGAAC 2553
333 rGlyThrAsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeu 350
|||||
2554 AGCAACACCAAGAAATTTGGCAAGGAAATGCTGACAGGCTCTATTGTTG 2603
350 1uLeuLeuGlyPheLysValProGlnArgGluProThrLysProAlaLeu 366
|||
2604 AACTGCTG.....GAAGCTTTCAAGTCAACACCAACCAACCAACACAG 2647
367 LysSerGluGlu..... 370
2648 TCATCGAGGAATGCTCACTTCGCCAACAAATGTCCCGCTTACAGCGCC 2697
371 .....LysThrPro 374
2698 CCGAGTGAAGCAACTGCGGAAGCAAGTCCCATGATGACCAACGCCAG 2747
374 1e.....LysLysProGly 378
2748 TGGGTCCCATGCGCAGCATTTAATACTGCGCCAGCAAGAAACACAGA 2797
379 AspGlyArg...LysValThrPhePheGluProGlySerGlyAspGln 394
|||
2798 AAGAGAGGAGATCAAAATGTAATTGTCAAGTCTAACGTGGAATCCAAAGA 2847
394 nGlyThrSerAsnLysGluAspGluPheArgMetProTYrLeuSerHis 411
|||
2848 GAGAGAGGCAATTAAGAGAGTGCGACCTGACCTGAGCAAGTAAGAAC 2897
411 1ng1nPheProAlaGlyLeuLeuProMetValProGluValAlaGlnAla 427
|||
2898 ACAGTCAACACTCTGCG.....GACAGCAGCAACAGCAGC 2932
428 ValGlyValSerGlnGlyHisHisThrLysAspPheThrArgAlaAla 444
|||
2933 ACCGGGATTCGCAAGCC.....ACAGAGCGGCGTAG 2964
444 oAsnProAlaLysAlaThrValThrAlaMetLeuAlaArgGluLeuLeu 461
|||
2965 TGAGAGTGCACTGAATCTCAACCGCGAGCAATYAG..... 3001
461 YrGlyTYrThrSerProThrAlaGluThrLysLeuLysAsnHisSer 477
|||
3002 ..AGTGTGTAGTACCAACAGCAGCAATGTTGAGCCCAACACGAGTGT 3049
478 SerGlyHisValProHisGlyProLeuThrArgProSer..... 490
|||
```

```

3050 AACACCATGCCGAGATAAAAAACACGAGAGTAGCAGCAACAGTAC 3099
491 .....GluGlnLeuAspTyrL 496
3100 GAGTAACACACAAAGCGCTGAGTGCACATGAGAGCAGCAGCTCTGTACC 3149
496 euserArgValGlnGlyPheGlnValGluTyrLysAspPheProLysAsn 512
3150 TTAGTAAACTGCTGAGACTTGTGAGTCAACTTCTCGGACTATCCGAAAGGC 3199
513 AsnLysAsnGluPheValSerLeuIleAsnCysSerSerGlnProProLe 529
3200 AATCACAACGAGTTCTGACCATGTGACATTGTCCACACATCCGCCGCA 3249
529 uIleSerHisGlyIleGlyLysAspValGluSerCysHisAspMetAla 546
3250 GATTGCCATGGCGCTGGCAAGAGCTCCGAGGAGTCCGACATGATGCTG 3299
546 lAleuAsnIleLeuLysLeuLeuSerGluLeu 556
3300 CAAGCAATGCCCTTGAAAAATCCTCAGCAAGCTG 3331

```





DR P-PSDB: Y83108.  
 XX Isolated mammalian or Caenorhabditis elegans staufen proteins useful  
 PT for treating retroviral infections especially human immunodeficiency  
 PT virus infections (HIV)  
 XX  
 PS Claim 4; Figure 1b; 96pp; English.  
 XX  
 CC Staufen is a RNA binding protein which interacts with double stranded  
 CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA  
 CC transport and localization. Mammalian SP contain multiple ds  
 CC RNA-binding domains. Chimeric proteins comprising portions of  
 CC staufen with proteins with RNase or protease activity can be used to  
 CC treat virus infections. The RNase or protease activity of the fusion  
 CC protein prevents proper maturation of the virus. Mammalian staufen  
 CC proteins seem to recognise double stranded RNA structure rather than  
 CC any sequence specific position. 2 bacterially-expressed fusion  
 CC proteins used in an RNA-binding assay (his/hsp (human SP) and  
 CC MBP/msp (murine SP)) both proteins strongly bound double stranded  
 CC RNA. Both fusion proteins also directly bound labeled ds RNAs and  
 CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the  
 CC treatment of RNA virus infections, especially those caused by  
 CC retroviruses, in particular human immunodeficiency virus. This  
 CC sequence is an alternatively spliced sequence to the one given in  
 CC Z9332.  
 XX  
 SQ Sequence 3142 BF; 862 A; 778 C; 733 G; 769 T; 0 other;

Query Match 99.9%; Score 3138.8; DB 21; Length 3142;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 3140; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 actctctgcccggctgaggcgctgagcgtcttcacgctttgagcgcgctgagcgt 60  
 DB 1 actctctgcccggctgaggcgctgagcgtcttcacgctttgagcgcgctgagcgt 60  
 QY 61 ctctctgcccggctgaggcgctgagcgtcttcacgctttgagcgcgctgagcgtc 120  
 DB 61 ctctctgcccggctgaggcgctgagcgtcttcacgctttgagcgcgctgagcgtc 120  
 QY 121 ctccacggcgaactcgcgtcttcctcctcctcctcctcctcctcctcctcctc 180  
 DB 121 ctccacggcgaactcgcgtcttcctcctcctcctcctcctcctcctcctcctc 180  
 QY 121 ctccacggcgaactcgcgtcttcctcctcctcctcctcctcctcctcctcctc 180  
 DB 121 ctccacggcgaactcgcgtcttcctcctcctcctcctcctcctcctcctcctc 180  
 QY 181 ctctctcctcctcctcgcgcgcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 240  
 DB 181 ctctctcctcctcctcgcgcgcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 240  
 QY 241 agcgcgcgaagaagataacccctactctgagaaactaaatgacatctgacatctgaa 300  
 DB 241 agcgcgcgaagaagataacccctactctgagaaactaaatgacatctgacatctgaa 300  
 QY 301 aaaaacacatcatcaagcctgtacacctactcctcgtatcagatccacataactaca 360  
 DB 301 aaaaacacatcatcaagcctgtacacctactcctcgtatcagatccacataactaca 360  
 QY 361 acatgagagagagtgctatcccccgcgagtaatttaccattccacgttccaccttac 420  
 DB 361 acatgagagagagtgctatcccccgcgagtaatttaccattccacgttccaccttac 420  
 QY 421 ttatcatagtgaaactctctgtgagagacagcaatctaagcgaagaagacaagac 480  
 DB 421 ttatcatagtgaaactctctgtgagagacagcaatctaagcgaagaagacaagac 480  
 QY 481 aggcctgcaaaacacgatactgtcgcacaaagcgttgaagatcctcagataagccctgc 540  
 DB 481 aggcctgcaaaacacgatactgtcgcacaaagcgttgaagatcctcagataagccctgc 540  
 QY 541 cagagagagcctgagagtgaaatgaaagaaatccgaagaagaatctcaataacttgaa 600  
 DB 541 cagagagagcctgagagtgaaatgaaagaaatccgaagaagaatctcaataacttgaa 600

QY 601 taagtcaagtgcttgagatgtcacttaaacggaacttgcctgtgaaatttcgaagtgagcc 660  
 DB 601 taagtcaagtgcttgagatgtcacttaaacggaacttgcctgtgaaatttcgaagtgagcc 660  
 QY 661 gggagagtgcccaaccacatagaagaacttctgaccaaaggttcgtgtggaggttg 720  
 DB 661 gggagagtgcccaaccacatagaagaacttctgaccaaaggttcgtgtggaggttg 720  
 QY 721 tggggggaagtggaaggggaagaagatcttcaaaaataatgccccttaactgttc 780  
 DB 721 tggggggaagtggaaggggaagaagatcttcaaaaataatgccccttaactgttc 780  
 QY 781 ttggagagctgaagaagtgtaaccgcccctgtcgtgagtgaaacgagtaagcctagaatca 840  
 DB 781 ttggagagctgaagaagtgtaaccgcccctgtcgtgagtgaaacgagtaagcctagaatca 840  
 QY 841 aaaaagaaaacaaacccaatagtcagaccacagaacgcccagatataatggccaagggatca 900  
 DB 841 aaaaagaaaacaaacccaatagtcagaccacagaacgcccagatataatggccaagggatca 900  
 QY 901 atccgattagccgagctgtgcccagatccagcagcgaagaaagagaagagccagatca 960  
 DB 901 atccgattagccgagctgtgcccagatccagcagcgaagaaagagaagagccagatca 960  
 QY 961 cgtctctcacagagcgaagcgtcccgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1020  
 DB 961 cgtctctcacagagcgaagcgtcccgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1020  
 QY 1021 gaaacacacatctgcagaagaaggaacgggcaccaaagaagtgagccaagcgaatcagcgcg 1080  
 DB 1021 gaaacacacatctgcagaagaaggaacgggcaccaaagaagtgagccaagcgaatcagcgcg 1080  
 QY 1081 agaacatgtctggaagatcctctgttcaaaagtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1140  
 DB 1081 agaacatgtctggaagatcctctgttcaaaagtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1140  
 QY 1141 tcaagtcagaagagaagacacccaataaagaacacagagtgagagaagaataaacttct 1200  
 DB 1141 tcaagtcagaagagaagacacccaataaagaacacagagtgagagaagaataaacttct 1200  
 QY 1201 ttgaacctgtctctgtggagtgaaaaatgggactaglaaataaagaagatgagttcagatgc 1260  
 DB 1201 ttgaacctgtctctgtggagtgaaaaatgggactaglaaataaagaagatgagttcagatgc 1260  
 QY 1261 ctatctaatgcatcagcagatctgtcgtctggaattcttccatgtgtgcccaggtgcgc 1320  
 DB 1261 ctatctaatgcatcagcagatctgtcgtctggaattcttccatgtgtgcccaggtgcgc 1320  
 QY 1321 aggcctgtagaggttagtcaagaagacatcaccaagaatttaccagagcgaatccgaatc 1380  
 DB 1321 aggcctgtagaggttagtcaagaagacatcaccaagaatttaccagagcgaatccgaatc 1380  
 QY 1381 ctgccaagggcacaagtgtaactgcataagcagagagtgctgtaatggggcactgcgc 1440  
 DB 1381 ctgccaagggcacaagtgtaactgcataagcagagagtgctgtaatggggcactgcgc 1440  
 QY 1441 ccacagccgagacatcttaagaataaacaatcttcaagcgaagtaaccataatgagactc 1500  
 DB 1441 ccacagccgagacatcttaagaataaacaatcttcaagcgaagtaaccataatgagactc 1500  
 QY 1501 tcaaggaagcctctggaacaaatggaatacttccagagtcacaggaattccaggttgat 1560  
 DB 1501 tcaaggaagcctctggaacaaatggaatacttccagagtcacaggaattccaggttgat 1560  
 QY 1561 acaagaacttccccaacaaacaaagaagaatctgtatcttcatcaatgtcctctc 1620  
 DB 1561 acaagaacttccccaacaaacaaagaagaatctgtatcttcatcaatgtcctctc 1620  
 QY 1621 agcgaactctgaatcagccaatgtatcggcaagagtgagagtccttcgaatgtatgtgctg 1680  
 DB 1621 agcgaactctgaatcagccaatgtatcggcaagagtgagagtccttcgaatgtatgtgctg 1680  
 QY 1681 cgtgaacatcttaagtgctgtctgagtgagccaacaaagtaacagagatgccaagaa 1740

Db	2761	tattaaaaaacatggtgatactataaataatgctctcttgaagccttagctttcacaact	2820
Qy	2821	acagagatagatcctcactgtagtgcataataaactgcagagtgaatttccagagtctcg	2880
Db	2821	acagagataatgctccctcactgtagtgcataataaactgcagagtgaatttccagagtctcg	2880
Qy	2881	atactgttaattacatctcatttaaggctcgaaaagaatgacttaagtttctgataacgc	2940
Db	2881	atactgttaattacatctcatttaaggctcgaaaagaatgacttaagtttctgataacgc	2940
Qy	2941	tgtgttccttttgatgtgtgtgttaccgtacacagaagtgctgcaactgaagctctgcgtg	3000
Db	2941	tgtgttccttttgatgtgtgtgttaccgtacacagaagtgctgcaactgaagctctgcgtg	3000
Qy	3001	tggtcgcgtatgaaaaaactcgttagccctcgagtgtaagtaactgccttcattcatgttta	3060
Db	3001	tggtcgcgtatgaaaaaactcgttagccctcgagtgtaagtaactgccttcattcatgttta	3060
Qy	3061	cgctcgtgaattttctcccatgtgaatgtaagtaaaacttaagtgtttgtcatcataaat	3120
Db	3061	cgctcgtgaattttctcccatgtgaatgtaagtaaaacttaagtgtttgtcatcataaat	3120
Qy	3121	ggtaatctaaataaaaaaaa 3142	
Db	3121	ggtaatctaaataaaaaaaa 3142	
RESULT 2			
ID	2933323	standard; cDNA; 3217 BP.	
XX	XX	2933323;	
AC	XX		
XX	XX		
DT	XX	04-JUL-2000 (first entry)	
XX	XX		
DE	XX	Human stauflen cDNA.	
XX	XX		
KW	XX	Staufen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;	
KW	XX	HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;	
KW	XX	acquired immune deficiency syndrome; protease; human; ss.	
OS	XX	Homo sapiens.	
XX	XX		
FH	XX	Location/Qualifiers	
FT	XX	CDS	363..1853
FT	XX		/+tag= a
FT	XX		/product= Staufen
XX	XX		
PN	XX	CA2238656-A1.	
XX	XX		
PD	XX	22-NOV-1999.	
XX	XX		
PE	XX	22-MAY-1998; 98CA-2238656.	
XX	XX		
PR	XX	22-MAY-1998; 98CA-2238656.	
XX	XX		
PA	XX	(UYMO-) UNIV MONTREAL.	
XX	XX		
PI	XX	Duchaine T, Wickham L, Cohen EA, Luo M, Desgroselliers L;	
PI	XX	Mouland AJ;	
DR	XX	WPI: 2000-246924/22.	
DR	XX	P-PSDB; Y83108.	
XX	XX		
PT	XX	Isolated mammalian or Caenorhabditis elegans staufen proteins useful	
PT	XX	for treating retroviral infections especially human immunodeficiency	
PT	XX	virus infections (HIV)	
XX	XX		
PS	XX	Claim 4; Figure 1a; 96pp; English.	
XX	XX		
CC	XX	Staufen is a RNA binding protein which interacts with double stranded	
CC	XX	RNA and/or rough endoplasmic reticulum. It is involved in mRNA	
CC	XX	transport and localization. Mammalian SP contain multiple ds	





QY 1786 ctggccatgaaccattataaaatcccaacatatatactgaaatactctgaactgttga 1845  
 Db 1861 ctggccatgaaccattataaaatcccaacatatatactgaaatactctgttga 1920  
 QY 1846 aaatttgaaattctgatacctccagtgccgagagacagtggtggaagatgttgg 1905  
 Db 1921 aaatttgaaattctgatacctccagtgccgagagacagtggtggaagatgttgg 1980  
 QY 1906 cagcagcaggaagaacacagaacacagaagagcgctgtgtccgctgtgactgtgtgc 1965  
 Db 1981 cagcagcaggaagaacacagaacacagaagagcgctgtgtccgctgtgactgtgtgc 2040  
 QY 1966 ggggttctgttgaatgagcaactcgtgtgacctggcgctccctacgcaatagcagctgt 2025  
 Db 2041 ggggttctgttgaatgagcaactcgtgtgacctggcgctccctacgcaatagcagctgt 2100  
 QY 2026 gggggaagaaggctccagcagctgttctcccggaacacagatccacacccctg 2085  
 Db 2101 gggggaagaaggctccagcagctgttctcccggaacacagatccacacccctg 2160  
 QY 2086 ggcacctcgtgttctgtcttcttccctgtgtgaaagaagaacgacagacccct 2145  
 Db 2161 ggcacctcgtgttctgtcttcttccctgtgtgaaagaagaacgacagacccct 2220  
 QY 2146 tctcaagctgtgtcactcagacacatctggagacaacccctgtgacagccatgcccagaagaag 2205  
 Db 2221 tctcaagctgtgtcactcagacacatctggagacaacccctgtgacagccatgcccagaagaag 2280  
 QY 2206 ggccttggacccgccccagagctcctaaagaacccagagaataatcgaatgtcttctcactcagcg 2265  
 Db 2281 ggccttggacccgccccagagctcctaaagaacccagagaataatcgaatgtcttctcactcagcg 2340  
 QY 2266 tgaaccaacttctcagtggtgagcagcgccccacccctcctcagtaaccacacacacac 2325  
 Db 2341 tgaaccaacttctcagtggtgagcagcgccccacccctcctcagtaaccacacacacacac 2400  
 QY 2326 cactgtcttctctcccaacagtatctgtatctagtgttcaatattcttcttgaatga 2385  
 Db 2401 cactgtcttctctcccaacagtatctgtatctagtgttcaatattcttcttgaatga 2460  
 QY 2386 tatggcacatataaaatttcaatttgagaatttctcaatgtatctagttaataatgacac 2445  
 Db 2461 tatggcacatataaaatttcaatttgagaatttctcaatgtatctagttaataatgacac 2520  
 QY 2446 agttggaactgtctgagcagcttatacaataatcaacacgaacagatcatatcc 2505  
 Db 2521 agttggaactgtctgagcagcttatacaataatcaacacgaacagatcatatcc 2580  
 QY 2506 atgtgtatgtgtgtgagacatttataatgactaacccagagcagttcagtgatgc 2565  
 Db 2581 atgtgtatgtgtgtgagacatttataatgactaacccagagcagttcagtgatgc 2640  
 QY 2566 aaattgtgtgccctgtgtcagctgaaacagtcctgtgacttccaaaacctggaatga 2625  
 Db 2641 aaattgtgtgccctgtgtcagctgaaacagtcctgtgacttccaaaacctggaatga 2700  
 QY 2626 tctcccaacagtgataaaatttgagaactttagaattttaaacttagatgttcatgtg 2685  
 Db 2701 tctcccaacagtgataaaatttgagaactttagaattttaaacttagatgttcatgtg 2760  
 QY 2686 ttcaattttatcatattttatttctgttaatgcaaacagagacttaattgaacttgtt 2745  
 Db 2761 ttcaattttatcatattttatttctgttaatgcaaacagagacttaattgaacttgtt 2820  
 QY 2746 ctctgttttaagaattataaaacattgtgtatctatatacatatgtgctctgagactt 2805  
 Db 2821 ctctgttttaagaattataaaacattgtgtatctatatacatatgtgctctgagactt 2880  
 QY 2806 agcttcaactaactacagagatataatctccatgtatgttccatataaactgtcagagtgat 2865  
 Db 2881 agcttcaactaactacagagatataatctccatgtatgttccatataaactgtcagagtgat 2940  
 QY 2866 ttccagagtgctcgatcgttaattatcatctcattagggctgaaagaatgacctac 2925

Db 2941 ttccagagtgctcgatcgttaattatcatctcattagggctgaaagaatgacctac 3000  
 QY 2926 gtcttcgtatagcgtgtgtcttcttgaatgtgtgttacttaacagaagtgtgtga 2985  
 Db 3001 gtcttcgtatagcgtgtgtcttcttgaatgtgtgttacttaacagaagtgtgtga 3060  
 QY 2986 ctgagcctcgt 3045  
 Db 3061 ctgagcctcgt 3120  
 QY 3046 tccatcaattgtttacgtcgtgaatttctcccatgtgaatgaatgaatgaatgt 3105  
 Db 3121 tccatcaattgtttacgtcgtgaatttctcccatgtgaatgaatgaatgaatgt 3180  
 QY 3106 ttgtcatcaataatgttaattacttaaaaaaa 3142  
 Db 3181 ttgtcatcaataatgttaattacttaaaaaaa 3217  
 RESULT 3  
 X90786  
 ID X90786 standard; DNA; 3190 BP.  
 XX  
 AC X90786;  
 XX  
 DT 13-JAN-2000 (first entry)  
 XX  
 DE DNA encoding human staufen protein.  
 XX  
 KW Human staufen protein; hStau; protein lysate; lung; kidney; testis;  
 KW ovary; brain; heart; telomerase RNA; complex; catalytic subunit;  
 KW telomerase assembly; maturation; transport; regulation; interaction;  
 KW inhibitory agent; antibody; immunosassy; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FH CDS 345..1835  
 FT /\*tag= a  
 FT /product= "Human staufen protein"  
 FT /note= "Binds to telomerase RNA"  
 FT polyA\_signal 3172..3177  
 FT /\*tag= b  
 XX  
 PN W09951255-A1.  
 XX  
 PD 14-OCT-1999.  
 XX  
 PF 06-APR-1999; 99WO-US07533.  
 XX  
 PR 06-APR-1998; 98US-0080783.  
 XX  
 PA (UYUO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.  
 XX  
 PI Greider CW, Le S;  
 DR WPI: 1999-620168/53.  
 DR P-PSDB: Y26335.  
 XX  
 PT Human staufen polypeptide useful in methods for identifying telomerase  
 PT inhibitors -  
 PS Claim 13; Page 41-42; 50pp; English.  
 XX  
 CC The present sequence is an isolated DNA encoding human staufen (hStau)  
 CC protein. The hStau protein was detected in protein lysates from lung,  
 CC kidney, testis and ovary, but not in brain and heart. It binds  
 CC specifically to human telomerase RNA and can complex with the  
 CC telomerase catalytic subunit. The hStau protein may play a role in  
 CC telomerase assembly, maturation, transport and regulation. Interaction  
 CC between hStau and telomerase can be inhibited by administering inhibitory  
 CC agents to the cell. These telomerase inhibitors can be identified using



QY	2020	gcccgtgvggaaagaagggctgcgcccagcagcgtgctctcccggaacacagcatccac	2079
Dp	2078	gcccgtgvggaaagaagggctgcgcccagcagcgtgctctcccggaacacagcatccac	2137
QY	2080	acccttggaaccccggttggtctctttttccctctgtggaagaaaggaacgacg	2139
Dp	2138	acccttggaaccccggttggtctctttttccctctgtggaagaaaggaacgacg	2197
QY	2140	accctctcaagctctgctcaacccagacacatctggaacaaacccctggacaacgacgacg	2199
Dp	2198	accctctcaagctctgctcaacccagacacacatctggaacaaacccctggacacgacgacg	2257
QY	2200	agagaagcctctgaacggcccgagcctaagaacacagagaanaaactaatgtctctac	2259
Dp	2258	agagaagcctctgaacggcccgagcctaagaacacagagaanaaactaatgtctctac	2317
QY	2260	tcgaacgtgacccaacactttctagctgacgaagggcccccacactctctgcatcccaac	2319
Dp	2318	tcgaacgtgacccaacactttctagctgacgaagggcccccacactctctgcatcccaac	2377
QY	2320	catcaacacgctctctctctccacaacgctgacgtatctcttctggtctcaattctcttt	2379
Dp	2378	catcaacacgctctctctctccacaacgctgacgtatctcttctggtctcaattctcttt	2437
QY	2380	gattgatatacacatataaanaatttcaatttgagaattctccaatgylatctagttaa	2439
Dp	2438	gattgatatacacatataaanaatttcaatttgagaattctccaatgylatctagttaa	2497
QY	2440	tagacaacgcttggaactgtctggaacacgactcttaataatactaacgcgaagaatc	2499
Dp	2498	tagacaacgcttggaactgtctggaacacgactcttaataatactaacgcgaagaatc	2557
QY	2500	atacacaatgctgaatggtgttagagcattttatcttaacctgacaacccggagaagttacg	2559
Dp	2558	atacacaatgctgaatggtgttagagcattttatcttaacctgacaacccggagaagttacg	2617
QY	2560	tgatgcaaatctgtggtgcctctgctcaagctgaanaagcctggaccttccaanaacctg	2619
Dp	2618	tgatgcaaatctgtggtgcctctgctcaagctgaanaagcctggaccttccaanaacctg	2677
QY	2620	aataagctctccacaagtctgataaattggaacaaattagaattttaaacttatagatgc	2679
Dp	2678	aataagctctccacaagtctgataaattggaacaaattagaattttaaacttatagatgc	2737
QY	2680	attggtgtccaattttatctcatctttatctttgttaagcaanaagacttaaaagac	2739
Dp	2738	attggtgtccaattttatctcatctttatctttgttaagcaanaagacttaaaagac	2797
QY	2740	tttgatctctgtttlaaagcttaataaaaaacatgctgatatatacatatggtcctgca	2799
Dp	2798	tttgatctctgtttlaaagcttaataaaaaacatgctgatatatacatatggtcctgca	2857
QY	2800	ggactatgcttccaactacaactacagatatgatctccatgtagtccatatataaacctgcag	2859
Dp	2858	ggactatgcttccaactacaactacagatatgatctccatgtagtccatatataaacctgcag	2917
QY	2860	agttatcttccaaggtgtctgatacgtttaactaacctccacttggtggtctgaagaagatg	2919
Dp	2918	agttatcttccaaggtgtctgatacgtttaactaacctccacttggtggtctgaagaagatg	2977
QY	2920	accctagcttctgataacagctgtctgctcttgatctgtgtctgtaactgtaacaaagtg	2979
Dp	2978	accctagcttctgataacagctgtctgctcttgatctgtgtctgtaactgtaacaaagtg	3037
QY	2980	ctgtgacatgaagcctctgcgtgtgtgtccgtatctgnaaaacctgtgtaacctgcagtgtaagt	3039
Dp	3038	ctgtgacatgaagcctctgcgtgtgtgtccgtatctgnaaaacctgtgtaacctgcagtgtaagt	3097
QY	3040	actgtctccatctaatgttttaagcctggaattttctccccaatggagatgaaataaactt	3099
Dp	3098	actgtctccatctaatgttttaagcctggaattttctccccaatggagatgaaataaactt	3157
QY	3100	aagctctgtcatcaataaatagttlaactataa	3132

Db	3158	aagtgcttcatacataatggttaacttcaaa	3190
<p>     </p>			
RESULT 4			
ID	Z93322	standard; cDNA; 3506 BP.	
XX	Z93322;		
AC	Z93322;		
XX	04-JUL-2000	(first entry)	
DT			
DE	Human staufen cDNA.		
XX			
KW	Staufen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;		
KW	HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;		
KW	acquired immune deficiency syndrome; protease; human; ss.		
XX			
OS	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	409..2142	
FT		//tag=a	
FT		/product= Staufen	
XX			
PN	CA2238656-A1.		
PD			
XX	22-NOV-1999.		
PF			
XX	22-MAY-1998; 98CA-2238656.		
PR			
XX	22-MAY-1998; 98CA-2238656.		
PA	(UYMO-) UNIV MONTREAL.		
XX			
PI	Duchaine T, Wickham L, Cohen EA, Luo M, Desrosiers L,		
PI	Mouland AJ;		
XX			
DR	WPI: 2000-246924/22.		
DK	P-PSDB: Y83023.		
PT	Isolated mammalian or Caenorhabditis elegans staufer proteins useful		
PT	for treating retroviral infections especially human immunodeficiency		
XX	virus infections (HIV)		
XX			
PS	Claim 4; Figure 1a; 96pp; English.		
XX			
CC	Staufen is a RNA binding protein which interacts with double stranded		
CC	RNA and/or rough endoplasmic reticulum. It is involved in mRNA		
CC	transport and localization. Mammalian sp contain multiple ds		
CC	RN-binding domains. Chimeric proteins comprising portions of		
CC	staufen with proteins with RNase or protease activity can be used to		
CC	treat virus infections. The RNase or protease activity of the fusion		
CC	protein prevents proper maturation of the virus. Mammalian staufen		
CC	proteins seem to recognise double stranded RNA structure rather than		
CC	any sequence specific position. 2 bacterially-expressed fusion		
CC	proteins used in an RNA-binding assay (his/hsp (human SP) and		
CC	MBP/msp (murine SP)) both proteins strongly bound double stranded		
CC	RNA. Both fusion proteins also directly bound labeled ds RNAs and		
CC	RNA/DNA hybrids. Staufen proteins may therefore be useful in the		
CC	treatment of RNA virus infections, especially those caused by		
CC	retroviruses, in particular human immunodeficiency virus.		
XX			
SQ	Sequence 3506 BP; 965 A; 875 C; 790 G; 876 T; 0 other:		
<p>Query Match 92.2%; Score 2896; DB 21; Length 3506; Best Local Similarity 100.0%; Pred. No. 0; Matches 2896; Conservative 0; Mismatches 0; Indels 0; Gaps 0;</p>			
OY	247	cagaagcacaaccctactgtagaactaatgcacgtgtcatgaacttggaaaaaac	306
<p>     </p>			
Db	611	cagaagcacaaccctactgtagaactaatgcacgtgtcatgaacttggaaaaaac	670

QY	307	caatgataaagcctgttgacccttacctctcgcgatgcagctccactataactaatacaatgata	366
Db	671	caatgataaagcctgttgacccttacctctcgcgatgcagctccactataactaatacaatgata	730
QY	367	gaggaagtgctctatcccccggaaggtacttttaccattccacgtttccaccttacttattc	426
Db	731	gaggaagtgctctatcccccggaaggtacttttaccattccacgtttccaccttacttattc	790
QY	427	aagctggaactctctgttgggggagcaagcaatttatatgtgcaaaagaaagaaagcaagacgctg	486
Db	791	aagctggaactctctgttgggggagcaagcaatttatatgtgcaaaagaaagaaagcaagacgctg	850
QY	487	cgaataacgatactgctgcgtccaaagcgttgagagatcccttcagaaatgagcccttcgcaaga	546
Db	851	cgaataacgatactgctgcgtccaaagcgttgagagatcccttcagaaatgagcccttcgcaaga	910
QY	547	ggctcggaagggtgaaatgggaaggaatccggaagaagaataatctcaataatctgaataaagtc	606
Db	911	ggctcggaagggtgaaatgggaaggaatccggaagaagaataatctcaataatctgaataaagtc	970
QY	607	aagctgttctgagatctgcaacttaaaacgggaactctgctctgaaattctgaagcttgagccgggaga	666
Db	971	aagctgttctgagatctgcaacttaaaacgggaactctgctctgaaattctgaagcttgagccgggaga	1030
QY	667	gtggcccaacccccacatagaaagaaactctgtgcaccaaggttcctggttgaggagtctgtggggg	726
Db	1031	gtggcccaacccccacatagaaagaaactctgtgcaccaaggttcctggttgaggagtctgtggggg	1090
QY	727	aagcttggaagggaataaagaataaagaaattccaagaataatgcgcgcatactgttcttgagg	786
Db	1091	aagcttggaagggaataaagaataaagaaattccaagaataatgcgcgcatactgttcttgagg	1150
QY	787	agctcgaaagaaagtcttaacgccccctgcctgcgcagcttgaaagagtaaaagcctatgatacaaaaaga	846
Db	1151	agctcgaaagaaagtcttaacgccccctgcctgcgcagcttgaaagagtaaaagcctatgatacaaaaaga	1210
QY	847	aaacaaaaccccatagatcaagccacagacaagccacgaataatgaccagggatatactccga	906
Db	1211	aaacaaaaccccatagatcaagccacagacaagccacgaataatgaccagggatatactccga	1270
QY	907	ctttagccgaccttgccccgataatccagacagggcaaaaaggaaggaagccaaagatataacggctcc	966
Db	1271	ctttagccgaccttgccccgataatccagacagggcaaaaaggaaggaagccaaagatataacggctcc	1330
QY	967	tcacaagagcgagagcctcccgccgcccgcagaggagcttcttgatgcaggttgaaagcttgtaaaacc	1026
Db	1331	tcacaagagcgagagcctcccgccgcccgcagaggagcttcttgatgcaggttgaaagcttgtaaaacc	1390
QY	1027	acactctcagaagaagaaacgggacacacaagaaggtgtgcgaagcgcaatactgcagccgagaaca	1086
Db	1391	acactctcagaagaagaaacgggacacacaagaaggtgtgcgaagcgcaatactgcagccgagaaca	1450
QY	1087	tgctcggagatctcttggtttttaaagttcccgacgagcgagcccaaacccgacttcaagt	1146
Db	1451	tgctcggagatctcttggtttttaaagttcccgacgagcgagcccaaacccgacttcaagt	1510
QY	1147	caggggggaagagacacccataaagaataacccaggggatactgaagaataaactctttttgaac	1206
Db	1511	caggggggaagagacacccataaagaataacccaggggatactgaagaataaactctttttgaac	1570
QY	1207	ctgagctcttgggagatgaataatctgagactcagtaataaagagatbaagtctcagatgtccttacc	1266
Db	1571	ctgagctcttgggagatgaataatctgagactcagtaataaagagatbaagtctcagatgtccttacc	1630
QY	1267	taagttcatcagcagctgcgtctgtgaaattcttcccatgtgtgcgcgaaggttcgccagagctg	1326
Db	1631	taagttcatcagcagctgcgtctgtgaaattcttcccatgtgtgtgcgcgaaggttcgccagagctg	1690
QY	1327	tagagagcttaataagaagatcatcaaaccaagaattttaaccagggcgagctccgaatactctgcga	1386
Db	1691	tagagagcttaataagaagatcatcaaaccaagaattttaaccagggcgagctccgaatactctgcga	1750

QY	1387	agcgcaacgctgaacgctgacatgtagcccgagagctgtgtgtatgtgggacacctctgcgccaag	1446
Db	1751	agggccaacgctaacacgccaatgataagcccgagagctgtgtgtatgtgggacacctctgcgccaag	1810
QY	1447	ccgagacacattttaaagaataaacctcccttcagggcaagtaacccatggaactcccaaga	1506
Db	1811	ccgagacacattttaaagaataaacctcccttcagggcaagtaacccatggaactcccaaga	1870
QY	1507	gaacccctgagcaacctggaactatcttccagagctccagggatcccaagcttgaaataaag	1566
Db	1871	gaacccctgagcaacctggaactatcttccagagctccagggatcccaagcttgaaataaag	1930
QY	1567	acttccccaaaaaacaagaagaagatcttgatctttatcaattgctctcttcagccaac	1626
Db	1931	acttccccaaaaaacaagaagaagatcttgatctttatcaattgctctcttcagccaac	1990
QY	1627	ctctgatacagcaatgagctatcgagcaagagatctggagacccctgcgcacatgatatgctgcgcga	1686
Db	1991	ctctgatacagcaatgagctatcgagcaagagatctggagacccctgcgcacatgatatgctgcgcga	2050
QY	1687	acatcttaaagcttgctgcttgagcttgagacaacaagaagtaacagagatgccaagaacaagaa	1746
Db	2051	acatcttaaagcttgctgcttgagcttgagacaacaagaagtaacagagatgccaagaacaagaa	2110
QY	1747	acggagcccaatgctcgtgtgtgtggagagtgctgaaacctttctcgtgcgatgacaatatataa	1806
Db	2111	acggagcccaatgctcgtgtgtgtggagagtgctgaaacctttctcgtgcgatgacaatatataa	2170
QY	1807	atcccaacaatatatacagcgaataaacctgaaactgctttgaaataattggaattctctgatacc	1866
Db	2171	atcccaacaatatatacagcgaataaacctgaaactgctttgaaataattggaattctctgatacc	2230
QY	1867	tcacagctggccgagagacaacggtgtggttaaaggaatgttgagcaagcaagcaaggaaacaacag	1926
Db	2231	tcacagctggccgagagacaacggtgtggttaaaggaatgttgagcaagcaagcaaggaaacaacag	2290
QY	1927	aaacacaagagagagcgctgtgtgagcccgctgagacctgctggtgtttgttgtatggccaact	1986
Db	2291	aaacacaagagagagcgctgtgtgagcccgctgagacctgctggtgtttgttgtatggccaact	2350
QY	1987	cggtaaaccttgagccgcttccctccacgaataagcagcgtccctgtgggaaagaagagcctgcgccaag	2046
Db	2351	cggtaaaccttgagccgcttccctccacgaataagcagcgtccctgtgggaaagaagagcctgcgccaag	2410
QY	2047	cagcctggtcttcccgggagacaacgaatcacacacccctgggacaactccgtgtgttgctct	2106
Db	2411	cagcctggtcttcccgggagacaacgaatcacacacccctgggacaactccgtgtgttgctct	2470
QY	2107	tttttccccctgtgtgaaagaagaagaacgagacaaccccttcccaagctcgtgctcacaatga	2166
Db	2471	tttttccccctgtgtgaaagaagaagaagaacgagacaaccccttcccaagctcgtgctcacaatga	2530
QY	2167	cacatttggagaaacaccccttgagacagcctgccaagaagagagcccttggaccggtcccaagc	2226
Db	2531	cacatttggagaaacaccccttgagacagcctgccaagaagagagcccttggaccggtcccaagc	2500
QY	2227	taaaagccacgagagaaaaatcaaaatgcttcttcaatcagcgtgacccaacttttctagtggt	2286
Db	2591	taaaagccacgagagaaaaatcaaaatgcttcttcaatcagcgtgacccaacttttctagtggt	2650
QY	2287	ccaacgcccccaacacacctctcgagtaacccaacacatcacacactgcttctcttccaacag	2346
Db	2651	ccaacgcccccaacacacctctcgagtaacccaacacatcacacactgcttctcttccaacag	2710
QY	2347	tgaatctgatatcttagttcatatatatttctcttggatgtgatatgacaactataataaattt	2406
Db	2711	tgaatctgatatcttagttcatatatatttctcttggatgtgatatgacaactataataaattt	2770
QY	2407	catcttgagaaatttccaatgatacttagttaaaatgacaagaatttggaaactgtgtgaga	2466
Db	2771	catcttgagaaatttccaatgatacttagttaaaatgacaagaatttggaaactgtgtgaga	2830
QY	2467	ctgaccttataataaacttcaacgacaagaatcatatccatctgtgactgtgtgttgaagactt	2526

```

Db 2831 ctgacttatacaataactaacgcaagacatcatcatctgtatgtgtagacatt 2890
Oy 2527 ttattattatgactaacccaagacagtttcaagtgatgcaaatbtygcctcgtgtc 2586
Db 2881 ttattattatgactaacccaagacagtttcaagtgatgcaaatbtygcctcgtgtc 2950
Oy 2587 agctgaacagctcctgagcttccaacacctgtataagctccacaggtgtataaatt 2646
Db 2951 agctgaacagctcctgagcttccaacacctgtataagctccacaggtgtataaatt 3010
Oy 2647 ggaacaattagaatttaaacatttagatgacatcttggtccatttatttattttt 2706
Db 3011 ggaacaattagaatttaaacatttagatgacatcttggtccatttatttattttt 3070
Oy 2707 attttgttaatgcaaaagagcttaaatgtaacttgatctctgttttaaaatattaa 2766
Db 3071 attttgttaatgcaaaagagcttaaatgtaacttgatctctgttttaaaatattaa 3130
Oy 2767 aaaaactgtgatacatatacatatgctcttgagagcttagcttcaactacacacaga 2826
Db 3131 aaaaactgtgatacatatacatatgctcttgagagcttagcttcaactacacacaga 3190
Oy 2827 tatgactccatgtagtccatataaaacctgcagagtgatttccagagtgctgcagatg 2886
Db 3191 tatgactccatgtagtccatataaaacctgcagagtgatttccagagtgctgcagatg 3250
Oy 2887 ttaatacatctccatcttgagctggtgaaagaagaatgacctaagttctgtataacagctgtgt 2946
Db 3251 ttaatacatctccatcttgagctggtgaaagaagaatgacctaagttctgtataacagctgtgt 3310
Oy 2947 gctttgtagtgtgttctacacagagtggtgcacagagctgcgtcgtgtgtgtc 3006
Db 3311 gctttgtagtgtgttctacacagagtggtgcacagagctgcgtcgtgtgtgtc 3370
Oy 3007 gtaaggaacacctgtgtagcctgcagtgtaagtaactgttccattcaattgtttaagctgtg 3066
Db 3371 gtaaggaacacctgtgtagcctgcagtgtaagtaactgttccattcaattgtttaagctgtg 3430
Oy 3067 aatttttccccatggaatgtaagtaaaactaagtgtttgtcatcaataaagttaatt 3126
Db 3431 aatttttccccatggaatgtaagtaaaactaagtgtttgtcatcaataaagttaatt 3490
Oy 3127 actaaaaaactaaaaa 3142
Db 3491 actaaaaaactaaaaa 3506

```

```

PF 22-MAY-1998; 98CA-2238656.
XX
XX 22-MAY-1998; 98CA-2238656.
XX
PA (UWMO-) UNIV MONTREAL.
PI Duchaine T, Wickham L, Cohen EA, Luo M, Desrosellers L;
PI Moulard AJ;
DR WPI: 2000-246924/22.
DR P-PSDB: Y83108.
XX
XX Isolated mammalian or Caenorhabditis elegans staufen proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)
PS Claim 4; Figure 1b; 96pp; English.
XX
XX Staufen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC staufen with proteins with RNase or protease activity can be used to
CC treat virus infections. The RNase or protease activity of the fusion
CC protein prevents proper maturation of the virus. Mammalian staufen
CC proteins seem to recognize double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hsp (human SP) and
CC MBP/MSP (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus. This
CC sequence is an alternatively spliced sequence to the one given in
CC 293322.
XX
SQ Sequence 3260 BP; 907 A; 795 C; 759 G; 799 T; 0 other:

```

Query Match 92.0%; Score 2890.2; DB 21; Length 3260;  
 Best Local Similarity 94.6%; Pred. No. 0;  
 Matches 3086; Conservative 0; Mismatches 53; Indels 124; Gaps 4;

```

Oy 1 actctctgcgggctgctggggcgctgagcgctcttcaagctttgcggcgctgcgct 60
Db 1 actctctgcgggctgctggggcgctgagcgctcttcaagctttgcggcgctgcgct 60
Oy 61 ctctctgcggctccgcttctcttgaccgctcccccgcggcgcgcgccgctc 120
Db 61 ctctctgcggctccgcttctcttgaccgctcccccgcggcgcgcgccgctc 120
Oy 121 ctccacggccactccgctcttccctccctctg----- 153
Db 121 ctccacggccactccgctcttccctccctctgagcagcagcagcagcaggtgagt 180
Oy 154 -----tcccttctctc 168
Db 181 ttgtctctgtgcccagctgagtgcaagtgcgatctgcgtccactgcgaactcca 240
Oy 169 cct-----tttctcttctctccct----- 192
Db 241 cctccagtgtaagatttcccaactcagcctcccgataagtgatgattacagagttt 300
Oy 193 -----cctgcgcgcacccgcccagcagcgcgcggcgaggagctcgag 229
Db 301 althaacacttaacctctagaactgaacaagaacacaltgttcttgaaacgcctct 360
Oy 240 cagcagcagaagaactaacccctactgtagaactaagcaactgtgcgtgaacttga 299
Db 361 tttaaaagaagaactaacccctactgtagaactaagcaactgtgcgtgaacttga 420
Oy 300 aaaaaaacatgtataagcctgttgacccttactctcgatgcagtcacataactac 359

```

Db 421 aaaaaaaccaatgataaagcctctgttgacccttactctcgaatgacgt---ccataactac 477  
 QY 360 aacatgagagagtgctatcccccagagctacttaccattccagttccaccctta 419  
 Db 478 aacatgagagagtgctatcccccagagctacttaccattccagttccaccctta 537  
 QY 420 ctatatacaagtctgaactctctgttggaagagacaacatttaagtcaaaagagaacaaga 479  
 Db 538 ctatatacaagtctgaactctctgttggaagagacaacatttaagtcaaaagagaacaaga 597  
 QY 480 cagcgtctggaacaacagatgctgctgcccagaagctgtgagatccctgcaagaatgagccctg 539  
 Db 598 cagcgtctggaacaacagatgctgctgcccagaagctgtgagatccctgcaagaatgagccctg 657  
 QY 540 ccaagaagagcctggaagtgaaatggaagagaatcccgagaagaagaatactcaataactctgaa 599  
 Db 658 ccaagaagagcctggaagtgaaatggaagagaatcccgagaagaagaatactcaataactctgaa 717  
 QY 600 ataagctcaagtgcttgagatctgcaacttaaacggaactgctctgtaattctcgaagtgagcc 659  
 Db 718 ataagctcaagtgcttgagatctgcaacttaaacggaactgctctgtaattctcgaagtgagcc 777  
 QY 660 cggagagagtggtcccaaccccaatgagaagaactttgtgaccagaaggttctgttgagagatctt 719  
 Db 778 cggagagagtggtcccaaccccaatgagaagaactttgtgaccagaaggttctgttgagagatctt 837  
 QY 720 gtggggggaaggtgaaaggggaaagcaagaagatttcaaaagaaatggccgcaatgacttct 779  
 Db 838 gtggggggaaggtgaaaggggaaagcaagaagatttcaaaagaaatggccgcaatgacttct 897  
 QY 780 ctctgagagagctgaaagaaatctacccgcccctgctctgcaagtctgaaagtaaaagccttgaatc 839  
 Db 898 ctctgagagagctgaaagaaatctacccgcccctgctctgcaagtctgaaagtaaaagccttgaatc 957  
 QY 840 aaaaaagaaaaaaacccaatagctcaagccacaagacaagcccaagaataatgagccaggggact 899  
 Db 958 aaaaaagaaaaaaacccaatagctcaagccacaagacaagcccaagaataatgagccaggggact 1017  
 QY 900 aatccgacttagccgactgcccagatccagcagggcaaaaagagaaagagccagagatc 959  
 Db 1018 aatccgacttagccgactgcccagatccagcagggcaaaaagagaaagagagccagagatc 1077  
 QY 960 aagctctcctcaagagagcgagagcctcccgagcgagagaggtctgtgtaatgcaagtgaaaggt 1019  
 Db 1078 aagctctcctcaagagagcgagagcctcccgagcgagagaggtctgtgtaatgcaagtgaaaggt 1137  
 QY 1020 ggaacaacacactctgcaagaagagcgagccacaagaaggtgagccagcgcaatgacgac 1079  
 Db 1138 ggaacaacacactctgcaagaagagcgagccacaagaaggtgagccagcgcaatgacgac 1197  
 QY 1080 gagaacaatgctgagagatccttggttttcaaaagttcccgagcgagccagccacaacaacccgca 1139  
 Db 1198 gagaacaatgctgagagatccttggttttcaaaagttcccgagcgagccagccacaacaacccgca 1257  
 QY 1140 ctcaagctcaagagagagagacccaataaagaaacgaagagagagtgagagaaagtaacctt 1199  
 Db 1258 ctcaagctcaagagagagagacccaataaagaaacgaagagagagtgagagaaagtaacctt 1317  
 QY 1200 ttctgaacctggtctgtggatgtaaaaatgagactagtaataaagagatgagttcagagtg 1259  
 Db 1318 ttctgaacctggtctgtggatgtaaaaatgagactagtaataaagagatgagttcagagtg 1377  
 QY 1260 cctataatcaagctcaag 1319  
 Db 1378 cctataatcaagctcaag 1437  
 QY 1320 cagagctgtagagagtagcaagaagagatcacacaaagaatttaccagagcgagctccgagat 1379  
 Db 1438 cagagctgtagagagtagcaagaagagatcacacaaagaatttaccagagcgagctccgagat 1497  
 QY 1380 cctgccaagcgcaacgtagaactgcaatgtagccgagagagagagagagagagagagagagagag 1439  
 Db 1498 cctgccaagcgcaacgtagaactgcaatgtagccgagagagagagagagagagagagagagagag 1557

QY 1440 cccacagccgagacccattttaaagataacatctcttcagggccacgttaacccatgagact 1499  
 Db 1558 cccacagccgagacccattttaaagataacatctcttcagggccacgttaacccatgagact 1617  
 QY 1500 ctacagagacccctctgagaaactggaactatcttccagaagtcagagagatcccaagtgtaa 1559  
 Db 1618 ctacagagacccctctgagaaactggaactatcttccagaagtcagagagatcccaagtgtaa 1677  
 QY 1560 tacaaagacttcccaaaaacaaag 1619  
 Db 1678 tacaaagacttcccaaaaacaaag 1737  
 QY 1620 cagccacccctctgatacagcagtgatcccgcaagagatgagagatccctgccaatgataagct 1679  
 Db 1738 cagccacccctctgatacagcagtgatcccgcaagagatgagagatccctgccaatgataagct 1797  
 QY 1680 gogctgaacatcttaaaagtctgctgctgagtgagtggaacaaagaatgacagagatgccaaga 1739  
 Db 1798 gogctgaacatcttaaaagtctgctgctgagtgagtggaacaaagaatgacagagatgccaaga 1857  
 QY 1740 acaggaagacggaacaaatgctgt 1799  
 Db 1858 acaggaagacggaacaaatgctgt 1917  
 QY 1800 ttataaaatcccaacataataactgaaataactgaaactgcttgaataatlttgaaatltc 1859  
 Db 1918 ttataaaatcccaacataataactgaaataactgaaactgcttgaataatlttgaaatltc 1977  
 QY 1860 tgatacctccagtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1919  
 Db 1978 tgatacctccagtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 2037  
 QY 1920 acaacagaaacaaag 1979  
 Db 2038 acaacagaaacaaag 2097  
 QY 1980 ggcacactggtggaacttgagctggtcccaagcaataagcaactgctgtggtggtggtggtggtggtggt 2039  
 Db 2098 ggcacactggtggaacttgagctggtcccaagcaataagcaactgctgtggtggtggtggtggtggtggt 2157  
 QY 2040 ggcacagcagctggttctcccgagacacagcaagatccacaacccctggtgagactccgtgtc 2099  
 Db 2158 ggcacagcagctggttctcccgagacacagcaagatccacaacccctggtgagactccgtgtc 2217  
 QY 2100 tggctcttttttcccccgtgtgtgaagaagaagacggacagcccttctcaagctggtc 2159  
 Db 2218 tggctcttttttcccccgtgtgtgaagaagaagacggacagcccttctcaagctggtc 2277  
 QY 2160 actcagacacatctgagacaacccctgagacagcagatgccaagagagagagagagagagagagagag 2219  
 Db 2278 actcagacacatctgagacaacccctgagacagcagatgccaagagagagagagagagagagagagag 2337  
 QY 2220 ccagagctaaagaagcaccagagagaaatcaaatgcttctactactcagctgagccaaacttctc 2279  
 Db 2338 ccagagctaaagaagcaccagagagaaatcaaatgcttctactactcagctgagccaaacttctc 2397  
 QY 2280 tagtgtgccaagggcccccacccctctctgcaataccacaacatcaacatcaacatcttctctct 2339  
 Db 2398 tagtgtgccaagggcccccacccctctctgcaataccacaacatcaacatcaacatcttctctct 2457  
 QY 2340 ccaacagtgatctgatatcttagtttcaatttcttctcttcttcttcttcttcttcttcttcttcttct 2399  
 Db 2458 ccaacagtgatctgatatcttagtttcaatttcttctcttcttcttcttcttcttcttcttcttcttct 2517  
 QY 2400 aaatttcatcttggaatlttctcaatgtglatctagtttaaaatagcaagcttggaactgtg 2459  
 Db 2518 aaatttcatcttggaatlttctcaatgtglatctagtttaaaatagcaagcttggaactgtg 2577  
 QY 2460 tctgagactgagcttataataaacttaacccgaagaagataatataatccatggtgtatgtgtgt 2519  
 Db 2578 tctgagactgagcttataataaacttaacccgaagaagataatataatccatggtgtatgtgtgt 2637









CC characterised by the expression of various cancer associated antigens.  
 CC The invention provides nucleic acid sequences and encoded polypeptides  
 CC which are cancer associated antigen precursors expressed in human breast  
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
 CC lung cancer.  
 XX  
 SQ Sequence 1127 BP; 349 A; 265 C; 307 G; 203 T; 3 other;

Query Match 35.28; Score 1107.2; DB 20; Length 1127;  
 Best Local Similarity 99.48; Pred. No. 1e-270; Indels 1; Gaps 1;  
 Matches 1120; Conservative 0; Mismatches 6;

QY 445 gagagacgaacttaattatgcaagaagaacacagagctgcgaacacatgctgctg 504  
 |||||||  
 Db 1 gagagacgaacttaattatgcaagaagaacacagagctgcgaacacatgctgctg 60  
 QY 505 ccaagcgttgagagatcctcagaaatgagccctgcagagagctgaggtgaaatgaa 564  
 |||||||  
 Db 61 ccaagcgttgagagatcctcagaaatgagccctgcagagagctgaggtgaaatgaa 120  
 QY 565 gagatccgaagaagaataatcctaataatcgaataatgaatgcaagttgttgaattgac 624  
 |||||||  
 Db 121 gagatccgaagaagaataatcctaataatcgaataatgaatgcaagttgttgaattgac 180  
 QY 625 ttaacaggaacttgctcgtgaatttcgaagtgcgcggaagatgcgcacccacacatga 684  
 |||||||  
 Db 181 ttaacaggaacttgctcgtgaatttcgaagtgcgcggaagatgcgcacccacacatga 240  
 QY 685 agaacttctgacccaaggttcggttgaggagttgttgagggaaggtgaagggaagaaga 744  
 |||||||  
 Db 241 agaacttctgacccaaggttcggttgaggagttgttgagggaaggtgaagggaagaaga 300  
 QY 745 agaagatttcaaaaagaataatccgcataagctgttcttgaagaactgaagaagtaccgc 804  
 |||||||  
 Db 301 agaagatttcaaaaagaataatccgcataagctgttcttgaagaactgaagaagtaccgc 360  
 QY 805 cccgcctcgtgagttgaacgaatgaagcctatgaatcaaaaagaacacacatagta 864  
 |||||||  
 Db 361 cccgcctcgtgagttgaacgaatgaagcctatgaatcaaaaagaacacacatagta 420  
 QY 865 agccacagaacagcccggaataatgagcagggatcaatccgattagccgactggccaca 924  
 |||||||  
 Db 421 agccacagaacagcccggaataatgagcagggatcaatccgattagccgactggccaca 480  
 QY 925 tccgcagcagcaaaaagaagaagagccagatatacagctcctcaacagaagagccccc 984  
 |||||||  
 Db 481 tccgcagcagcaaaaagaagaagagccagatatacagctcctcaacagaagagccccc 540  
 QY 985 cgcgcgcagaggaattctgtgagtgaaagtgtgaaacacacatgcagaagaacag 1044  
 |||||||  
 Db 541 cgcgcgcagaggaattctgtgagtgaaagtgtgaaacacacatgcagaagaacag 600  
 QY 1045 gcacacaagaaggtgagcgaatgagcgcgaagaacatgctggaatctctggtt 1104  
 |||||||  
 Db 601 gcacacaagaaggtgagcgaatgagcgcgaagaacatgctggaatctctggtt 660  
 QY 1105 tcaagttccgcagcagcagcccaacacccgactcaatgaagaggaagaacaccca 1164  
 |||||||  
 Db 661 tcaagttccgcagcagcagcccaacacccgactcaatgaagaggaagaacaccca 720  
 QY 1165 taaagaacccggagtggaagaagaatlaaccttlttgaacccgtgctctggatgaa 1224  
 |||||||  
 Db 721 taaagaacccggagtggaagaagaatlaaccttlttgaacccgtgctctggatgaa 780  
 QY 1225 atggagactagtaataaagagagatgagtcagagatccttatcaatgaatcagaagctgc 1284  
 |||||||  
 Db 781 atggagactagtaataaagagagatgagtcagagatccttatcaatgaatcagaagctgc 840  
 QY 1285 ctgctggaattcttccatagtgccgaggtgcgcagagctgtagaggttagtcaagagac 1344  
 |||||||  
 Db 841 ctgctggaattcttccatagtggtgcccagagtgccagagctgtagaggttagtcaagagac 900

QY 1345 atcacaccaagaatttaccagggcagctccgaatcctgcgaagccaggttaactgcc 1404  
 |||||||  
 Db 901 atcacaccaagaatttaccagggcagctccgaatcctgcgaagccaggttaactgcc 960  
 QY 1405 tgatagcccgagagttgtgtatggtgggacactgcgcacag-ccgagacattttaaag 1463  
 |||||||  
 Db 961 tgatagcccgagagttgtgtatggtgggacactgcgcacagccgagacattttaaag 1020  
 QY 1464 aataacatcttccagcagcagctaccccatggaactctcagagaccccttgagacactg 1523  
 |||||||  
 Db 1021 aataacatcttccagcagcagctaccccatggaactctcagagaccccttgagacactg 1080  
 QY 1524 gactatcttccagagatccagagatccaggttgaatacaagaact 1570  
 |||||||  
 Db 1081 gactatcttccagagatccagagatccaggttgaatacaagaact 1127

RESULT 8  
 ID X40002  
 XX X40002 standard; DNA; 946 BP.  
 AC X40002;  
 XX  
 DT 02-JUL-1999 (first entry)  
 DE  
 DE Prostate cancer associated gene.  
 KW Cancer associated antigen; diagnosis; research; treatment; human;  
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
 KW prostate cancer; ss.  
 OS Homo sapiens.  
 XX  
 PN W09904265-A2.  
 XX  
 PD 28-JAN-1999.  
 XX  
 XX  
 PE 15-JUL-1998; 98W0-US14679.  
 PR 22-JUN-1998; 98US-0102322.  
 PR 17-JUL-1997; 97US-0896164.  
 PR 10-OCT-1997; 97US-0061599.  
 PR 10-OCT-1997; 97US-0061765.  
 PR 10-OCT-1997; 97US-0948705.  
 PR 11-OCT-1997; 97GB-0021697.  
 XX  
 PA (LUDWIG-) LUDWIG INST CANCER RES.  
 PI  
 PI Chen Y, Gout I, Gure A, O'Hare M, Ohta Y, Old LJ;  
 PI Pfeundschnuh M, Sahlin U, Scanlan MJ, Stockert E;  
 PI Tureci O;  
 XX  
 DR WPI: 1999-132448/11.  
 XX  
 XX

New isolated cancer associated nucleic acids and polypeptides -  
 PT isolated using sera from cancer patients, used to develop products  
 PT for the diagnosis, monitoring or treatment of cancers  
 XX  
 XX  
 PS Claim 67; Page 629-630; 787pp; English.

The invention relates to a method for diagnosing a disorder characterised  
 CC by expression of a human cancer associated antigen precursor coded for by  
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
 CC biological sample isolated from a subject with an agent that specifically  
 CC binds to the NAM, an expression product or a fragment of an expression  
 CC product complexed with an HLA molecule; and (b) determining the  
 CC interaction between the agent and the NAM or the expression product as a  
 CC determination of the disorder. The products and methods can be used in  
 CC the diagnosis, monitoring, research, or treatment of conditions  
 CC characterised by the expression of various cancer associated antigens.  
 CC The invention provides nucleic acid sequences and encoded polypeptides  
 CC which are cancer associated antigen precursors expressed in human breast  
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and

CC Lung cancer.  
 XX  
 SQ Sequence 946 BP; 292 A; 208 C; 259 G; 170 T; 17 other;

Query Match 25.4%; Score 798.8; DB 20; Length 946;  
 Best Local Similarity 93.6%; Pred. No. 1.2e-192;  
 Matches 888; Conservative 0; Mismatches 54; Indels 7; Gaps 6;

QY 445 gaggaagcaatcttaatgcaagaagaagacagagctgcgaaacacgattgctg 504  
 |||||  
 DB 1 gaggaagcaatcttaatgcaagaagaagacagagctgcgaaacacgattgctg 60  
 |||||  
 QY 505 ccaaacgttgaagatccctcagaatgagccctgcagagagagctggaagtgaatgaa 564  
 |||||  
 DB 61 ccaaacgttgaagatccctcagaatgagccctgcagagagagctggaagtgaatgaa 120  
 |||||  
 QY 565 gagaatccgaagaagaataatcacaataatctgaataaagtcgaatgattgac 624  
 |||||  
 DB 121 gagaatccgaagaagaataatcacaataatctgaataaagtcgaatgattgac 180  
 |||||  
 QY 625 ttaaaaggaacttgctgtgaatttcgaagtgccggggaagtgccacacatga 684  
 |||||  
 DB 181 ttaaaaggaacttgctgtgaatttcgaagtgccggggaagtgccacacatga 240  
 |||||  
 QY 685 agaaacttgagcaagaagttcgtgtgagagtggtgaggaagtggaaggaagaagca 744  
 |||||  
 DB 241 agaaacttgagcaagaagttcgtgtgagagtggtgaggaagtggaaggaagaagca 300  
 |||||  
 QY 745 agaaagatcacaagaagaataatgcccataagctgttcttgaagagctgaagaatlcaccg 804  
 |||||  
 DB 301 agaaagatcacaagaagaataatgcccataagctgttcttgaagagctgaagaatlcaccg 360  
 |||||  
 QY 805 ccttgctgtgagtgtaaacgagtaaacctagaatcacaagaagaagaacacacatgtca 864  
 |||||  
 DB 361 ccttgctgtgagtgtaaacgagtaaacctagaatcacaagaagaagaacacacatgtca 420  
 |||||  
 QY 865 agcacagaagaagccagaaatagtcgaagaggaatcaatccgattagccgctgcccaga 924  
 |||||  
 DB 421 agcacagaagaagccagaaatagtcgaagaggaatcaatccgattagccgctgcccaga 480  
 |||||  
 QY 925 tccagcagcacaagaagaaggaagcagacagctcctcacaagcgaagcctcc 984  
 |||||  
 DB 481 tccagcagcacaagaagaaggaagcagacagctcctcacaagcgaagcctcc 540  
 |||||  
 QY 985 cggcgccgagggaggttgatgacagtgtaaggttgaaaccacac-tgaagaaggaagc 1043  
 |||||  
 DB 541 cggcgccgagggaggttgatgacagtgtaaggttgaaaccacac-tgcaagaaggaagc 600  
 |||||  
 QY 1044 ggcacacaagaagaagtgagcgaagcgaatgacgagcagaacaatgctgagatcccttgt 1103  
 |||||  
 DB 601 ggcacacaagaagaagtgagcgaagcgaatgacgagcagaacaatgctgagatcccttgt 660  
 |||||  
 QY 1104 ttc-aaagtcgcgagcagcagccacacacacacacacacacacacacacacacac 1162  
 |||||  
 DB 661 ttc-aaagtcgcgagcagcagccacacacacacacacacacacacacacacacac 720  
 |||||  
 QY 1163 cataaagaacccgaggtgaggaagaagaagtaaacctttttgaaacccgctgctgggatga 1222  
 |||||  
 DB 721 attaagaagaacccgaggtgaggaagaagaagtaaac-ntttgaaacccgctgctgggatga 778  
 |||||  
 QY 1223 aaatggagactgaatcaaaagaagaatgattcagatgacttaataatcagcagact 1282  
 |||||  
 DB 779 aaatggg-cttgytaataaagaagaatgattcagatgacttaataatcagcagact-ct 836  
 |||||  
 QY 1283 gacctgtgaatctctcccatggtgcccgaagtcgcccagagctgtgaagttaagcaag 1342  
 |||||  
 DB 837 tgtgtgtggaatctctcccatggtgcccgaagtcgcccagagctgtgaagtta-ctang 895  
 |||||  
 QY 1343 acatcacacaagaatcttaaccagagcagctccgaatcctgccaagcc 1391  
 |||||  
 DB 896 ccattccacccaagaatcttcccgagcagtttccatctgccaagccc 944  
 |||||

RESULT 9  
 ID X40000  
 AC X40000 standard; DNA; 844 BP.  
 XX  
 AC X40000;  
 XX  
 DT 02-JUL-1999 (first entry)  
 XX

DE Prostate cancer associated gene.  
 XX

KW Cancer associated antigen; diagnosis; research; treatment; human;  
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
 KW prostate cancer; ss.  
 XX

OS Homo sapiens.  
 XX

PN W09904265-A2.  
 XX

PD 28-JAN-1999.  
 XX

PF 15-JUL-1998: 98WO-US14679.  
 XX

PR 22-JUN-1998: 98US-0102322.  
 XX

PR 17-JUL-1997: 97US-0896164.  
 XX

PR 10-OCT-1997: 97US-0061599.  
 XX

PR 10-OCT-1997: 97US-0061765.  
 XX

PR 10-OCT-1997: 97US-0948705.  
 XX

PR 11-OCT-1997: 97GB-0021697.  
 XX

PA (LUDW-) LUDWIG INST CANCER RES.  
 XX

PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;  
 XX

PI Pfreundschuh M, Sahlin U, Scanlan MJ, Stockert E;  
 XX

PI Tureci O;  
 XX

DR WPI: 1999-132448/11.  
 XX

PT New isolated cancer associated nucleic acids and polypeptides -  
 XX

PT Isolated using sera from cancer patients, used to develop products  
 XX

PT for the diagnosis, monitoring or treatment of cancers  
 XX

PS Claim 67: Page 629: 787pp: English.  
 XX

CC The invention relates to a method for diagnosing a disorder characterised  
 CC by expression of a human cancer associated antigen precursor coded for by  
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
 CC biological sample isolated from a subject with an agent that specifically  
 CC binds to the NAM, an expression product or a fragment of an expression  
 CC product complexed with an HLA molecule; and (b) determining the  
 CC interaction between the agent and the NAM or the expression product as a  
 CC determination of the disorder. The products and methods can be used in  
 CC the diagnosis, monitoring, research, or treatment of conditions  
 CC characterised by the expression of various cancer associated antigens.  
 CC The invention provides nucleic acid sequences and encoded polypeptides  
 CC which are cancer associated antigen precursors expressed in human breast  
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
 CC lung cancer.  
 CC

SQ Sequence 844 BP; 251 A; 196 C; 234 G; 156 T; 7 other;

Query Match 24.0%; Score 755.6; DB 20; Length 844;  
 Best Local Similarity 96.3%; Pred. No. 1e-181;

Matches 812; Conservative 0; Mismatches 26; Indels 5; Gaps 4;

QY 607 aagttgtgagattgacttaaaaggaacttgcttgaattcgaagtgccgggaga 666  
 |||||

DB 1 aagttgtgagattgacttaaaaggaacttgcttgaattcgaagtgccgggaga 60  
 |||||

QY 667 gtggccacccacatgaagaactttgacccaaggttcggtgaggttggtggg 726  
 |||||

DB 61 gtggccacccacatgaagaactttgacccaaggttcggtgaggttggtggg 120  
 |||||

OY	727	aagtgatgaagggaataagcaagaagatttccaagaanaaatgccgcataagctgttcttgagg	786
Db	121	aagggtgaagggaataaagcaagaagatttccaagaanaaatgcccgcataagctgttcttgagg	180
OY	787	agcttaagaanaatttaccgccccttgctctgacttgaaacggatgaagagcttgaaatcaaaaaga	846
Db	181	agctgaagaagaatttaccgccccttgctctgacttgaaacggatgaagagcttgaaatcaaaaaga	240
OY	847	aaacaaaaccccatagttcaagccacacagaaaccccagaatatgycacaggggagatcaatccga	906
Db	241	aaacaaaaccccatagttcaagccacacagaaaccccagaatatgycacaggggagatcaatccga	300
OY	907	tttagcgcagcttgccccagatccagcgcgcgcaaaaaagagaagaagaccagaatcacagctcc	966
Db	301	tttagcgcagcttgccccagatccagcgcgcgcaaaaaagagaagaagaccagaatcacagctcc	360
OY	967	tcaacagacagagctcccgccgcgcgcagagagttgtgatgatgaggttgaaagttgaaacc	1026
Db	361	tcaacagacagagctcccgccgcgcgcagagagttgtgatgatgaggttgaaagttgaaacc	420
OY	1027	acaatcgsaagaagaaacgggcaccaaacaagaagtgtgcgaacgcacatgcacgtcgagaaca	1086
Db	421	acaatcgsaagaagaaacgggcaccaaacaagaagtgtgcgaacgcacatgcacgtcgagaaca	480
OY	1087	tgtctgaatattcttgttttccaagtcccgccagcgacagccacacacccgcactccaagt	1146
Db	481	tgtctgaatattcttgttttccaagtcccgccagcgacagccacacacccgcactccaagt	540
OY	1147	cagaagagaagaacaccccaataagaagaacagaggatgtgagaagaagaatcaaccttttgaac	1206
Db	541	cagaagagaagaacaccccaataagaagaacagaggatgtgagaagaagaatcaaccttttgaac	600
OY	1207	ctgctctctgaggatgaaaaatgaggactagtaataaagaagatgagttcagatgctctatc	1266
Db	601	ctgctctctgaggatgaaaaatgaggactagtaataaagaagatgagttcagatgctctatc	660
OY	1267	taagatcatcagagatgctgctgtctgtaattcttcccatggtgcgcgaagtgcgccagctg	1326
Db	661	taagatcatcagagatgctgctgtctgtaattcttcccatggtgcgcgaagtgcgccagctg	720
OY	1327	tagg-agtttagtcaaggagcatcacaccocaaagaatttaccagggcagctccgaaatcctgcc	1385
Db	721	taggagatgttagtcaaggagcatcacaccocaaagaatttaccagggcagctccgaaatcctgcc	779
OY	1386	aaggccaagttacatgcatgataagcccagagagttgtgtatgtyggggcaccctgcgccaca	1445
Db	780	angmcmng--tactgcacatgataggcc-anagttgttgtatgtgggcancttgccca	836
OY	1446	gcc 1448	
Db	837	ggc 839	
RESULT 10			
X39992			
ID	X39992 standard; DNA: 773 BP.		
XX	X39992:		
XX			
XX	02-JUL-1999 (first entry)		
XX			
DE	Prostate cancer associated gene.		
XX			
KX	Cancer associated antigen; diagnosis; research; treatment; human;		
KW	breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;		
KW	prostate cancer; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	W09904265-A2.		
XX			
PD	28-JAN-1999.		

Query Match	Best Local Similarity	Matches	Score	DB	Length
24.0%;	99.1%;	0;	753.2;	20;	773;
Conservative	Pred. No. 3.9e-181;	Mismatches 6;	Indels 1;	Gaps 1;	
799	taccgcccctgcctcagttgaacgaagtaaacgtacgaatacaaaagaaacaaaccca	858			
1	taccgcccctgcctcagttgaacgaagtaaacgtacgaatacaaaagaaacaaaccca	60			
859	tagtaagcccaagcaaaccccaagaataagccaggggataatccgattagccagctgg	918			
61	tagtaagcccaagcaaaccccaagaataagccaggggataatccgattagccagctgg	120			
919	ccagatccagcagagcaaaaaggaaggaagcagagatfacacgcctccatcaagacgag	978			
121	ccagatccagcagagcaaaaaggaaggaagcagagatfacacgcctccatcaagacgag	180			
979	gctcccgccgcgcagggaggtttgtgatgcaggtgaaaggttgagaaaccaactgcagaag	1038			
181	gctcccgccgcgcagggaggtttgtgatgcaggtgaaaggttgagaaaccaactgcagaag	240			
1039	gaacgggaccccaagaaggtgtgccaagcgaatgtgcagccgaagaacatgctggagatcc	1098			
241	gaacgggaccccaagaaggtgtgccaagcgaatgtgcagccgaagaacatgctggagatcc	300			
1099	ttggtttcaaaatcccgacagcgcagcccaacaaacccgcacatccaagtacagagaaaga	1158			
301	ttggtttcaaaatcccgacagcgcagcccaacaaacccgcacatccaagtacagagaaaga	360			
1159	caaccataaagaaccaggggagtgaagaanaagtaacctttttgaacctggtcttggg	1218			
361	caaccataaagaaccaggggagtgaagaanaagtaacctttttgaacctggtcttggg	420			

```

OY 1219 atgaaaatggaactaataaagagatgagtcagatgcttataatgaatcacc 1278
DB 421 atgaaaatggaactaataaagagatgagtcagatgcttataatgaatcacc 480
OY 1279 agctgcctgctgaatctctccatgagtcgcccagagtcgcccagctgtgagagtc 1338
DB 481 agctgcctgctgaatctctccatgagtcgcccagagtcgcccagctgtgagagtc 540
OY 1339 aaggacataccacaaagtttaccagggagctcgatctctgccaagggcagagtaa 1398
DB 541 aaggacataccacaaagtttaccagggagctcgatctctgccaagggcagagtaa 600
OY 1399 ctgcctatgataccgagagttgttatgaggagcaccctgcaccaag -ccagagaccatt 1457
DB 601 ctgcctatgataccgagagttgttatgaggagcaccctgcaccaagcccgagaccatt 660
OY 1458 ttaagaataacatctcttccagagcagctacccatgagctctcaagagaccctctgag 1517
DB 661 ttaagaataacatctcttccagagcagctacccatgagctctcaagagaccctctgag 720
OY 1518 caactgactatcttccagagcagagcttccaggttgaatacaagaact 1570
DB 721 caactgactatcttccagagcagagcttccaggttgaatacaagaact 773

RESULT 11
X40001/c
ID X40001 standard; DNA; 825 BP.
XX
AC X40001;
XX
DT 02-JUL-1999 (first entry)
XX
DE Prostate cancer associated gene.
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0103322.
XX
PR 17-JUL-1997; 97US-0896164.
XX
PR 10-OCT-1997; 97US-0061599.
XX
PR 10-OCT-1997; 97US-0061765.
XX
PR 10-OCT-1997; 97US-0948705.
XX
PR 11-OCT-1997; 97GB-0021697.
XX
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ,
PI Pfeundschnh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI, 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Claim 67; Page 629; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically

```

```

CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
SQ Sequence 825 BP; 289 A; 152 C; 151 G; 228 T; 5 other;

Query Match          17.5%; Score 551; DB 20; Length 825;
Best Local Similarity 95.9%; Pred. No. 6.7e-130;
Matches 585; Conservative 0; Mismatches 23; Indels 2; Gaps 2;

OY 2525 ttattatcattgactaaaccagagacgttccagatgacaaatgtgtgccccttgg 2584
DB 791 TTTAATTTCTCCTTGCTAAACCCAGNCAGTTCAGTGAATGCAATTTGGGGGCCNTCTGGT 732
OY 2585 tcaagctgaaa -cagtcctgagcttcaaaaaacctgaataagctccacagttgataa 2643
DB 731 TCAGCTGAACACCGAGCTCTGAGACTTCAAAAACCTGAATAAGCTCCCGAGTTGATAA 672
OY 2644 attggaacatttagagatlttaactttagatgatcatcttggtccattttatccat 2703
DB 671 ATTGAGACATTTAGGAATTTTAACTTTAGATGATATTTGGCCATTTTATATNTTCATT 612
OY 2704 ttattcttctgtaagcaaacagagac -ttaaatagacttgaactgtctgttttaagaat 2762
DB 611 TTTATTTTGTATATGCAAAACAGAGACTTTAAATGAGACTTTGATCTCGTTTAAAGATTA 552
OY 2763 ttaaaaaaacattgtgtatctatcatatagatgctcttgaagagcttaagcttccactac 2822
DB 551 TTTAAAAACATTTGTATCTATATACATATGAGCTCTTGAAGCACTTATTCACACTAC 492
OY 2823 aggatagatctcagatgtagtcataataaacctgcagagatgtttccagagtgctcgat 2882
DB 491 AGGATATGATCTCCATGATGTAGTCATATTAACCTGCAGAGTGATTTTCCAGAGTCTCGAT 432
OY 2883 actgttaattacatctccatlaagagctgaagaagaatgacctagcttctgtatagctg 2942
DB 431 ACTGTTAATTACATCTCCATTTAGGCTGAAGAAAGATGACCTAGCTTGTATATACAGCTG 372
OY 2943 tgtgtcttctgagtggtgtactgtacacagagagtggtgacactgaagctctggtgtg 3002
DB 371 TGTGTCTTTTGATGTGTGTACTGTACACAGAGTGTGTGCACTGAGGCTGTGCGTGTG 312
OY 3003 gtccgatagaaacctgtgtagccctgcgagagttgaagtactgtctccatcatgttctag 3062
DB 311 GTCCGATATGGAAGAGCTGTGAGCCCTGCGAGTTAAGTACTGCTTCCATTCATTGTTTACG 252
OY 3063 ctggaatttctccccaatgaaatgaaatgaaactaaagtgtgtgttctcaataaagtg 3122
DB 251 CTTGGAATTTTTCCTCCCATGAGATGTAAAGTAAACTTAAGTGTGTATCAATTAATG 192
OY 3123 taactactaa 3132
DB 191 TAATACTAAA 182

RESULT 12
X39993/c
ID X39993 standard; DNA; 794 BP.
XX
AC X39993;
XX
DT 02-JUL-1999 (first entry)
XX
DE Prostate cancer associated gene.
XX

```

KM Cancer associated antigen: diagnosis; research; treatment; human;  
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
 KM prostate cancer; ss.  
 OS Homo sapiens.  
 XX  
 XX M09904265-A2.  
 XX  
 XX  
 PD 28-JAN-1999.  
 XX  
 PF 15-JUL-1998; 98WO-US14679.  
 XX  
 XX 22-JUN-1998; 98US-0102322.  
 PR 17-JUL-1997; 97US-0896164.  
 PR 10-OCT-1997; 97US-0061599.  
 PR 10-OCT-1997; 97US-0061765.  
 PR 10-OCT-1997; 97US-0048705.  
 PR 11-OCT-1997; 97GB-0021697.  
 XX  
 XX (LUDW-) LUDWIG INST CANCER RES.  
 XX  
 PI Chen Y, Gout I, Gure A, O'Hare M, Ohta Y, Old LJ;  
 PI Pfeundschnuh M, Sahin U, Scanlan MJ, Stockert E;  
 PI Tureci O;  
 XX  
 DR WPI: 1999-132448/11.  
 XX  
 XX New isolated cancer associated nucleic acids and polypeptides -  
 PT isolated using sera from cancer patients, used to develop products  
 PT for the diagnosis, monitoring or treatment of cancers  
 PS  
 PS Claim 67: Page 626; 787Pp: English.  
 XX  
 CC The invention relates to a method for diagnosing a disorder characterised  
 CC by expression of a human cancer associated antigen precursor coded for by  
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
 CC biological sample isolated from a subject with an agent that specifically  
 CC binds to the NAM, an expression product or a fragment of an expression  
 CC product complexed with an HLA molecule; and (b) determining the  
 CC interaction between the agent and the NAM or the expression product as a  
 CC determination of the disorder. The products and methods can be used in  
 CC the diagnosis, monitoring, research, or treatment of conditions  
 CC characterised by the expression of various cancer associated antigens.  
 CC The invention provides nucleic acid sequences and encoded polypeptides  
 CC which are cancer associated antigen precursors expressed in human breast  
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
 CC lung cancer.  
 CC  
 CC Sequence 794 BP: 273 A; 141 C; 152 G; 206 T; 22 other;  
 SQ

Query Match 16.0%; Score 501.4; DB 20; Length 794;  
 Best Local Similarity 96.3%; Pred. No. 2.4e-117;  
 Matches 517; Conservative 0; Mismatches 18; Indels 2; Gaps 1;

QY 2598 tcttggacttcaaaacctgaagtcctcc--acagtgtataaattgacacatt 2655  
 DB 718 TCTCTGACTTTCAAAACCTTGAATAGTCTCCCNCCAGTTGNAATTTGGCCNATT 659  
 QY 2656 aggaatttaaaccttagatgacatctgtgtcatttatttatttatttattt 2715  
 DB 658 AGGAAATTTAAACNTTAATGATGTTGTTCCATTTTTATTTATTTATTTTGGT 599  
 QY 2716 aatgaagaacgacttaaatgacttgcctctgttttaagaatataaacaatg 2775  
 DB 598 AATGCAAAACAGNCTTAAATGAACCTTGATCTCTTTTAAAGATTATTTAAACATTG 539  
 QY 2776 tttatcatatcatatgctctgaagacttgccttcaactcaacacagatgatctc 2835  
 DB 538 TGNATCTATACATATGCGCTTGGAGACTTACCTTACATACACAGATATATCTC 479  
 QY 2836 catgtatgcataataaacctgcagagtgatttcagagtgctgacatcgttaattaca 2895  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 478 CATGTAGTCATATATACCTGCAGATGATTTTCCANMGTCGATCTGTAAATACA 419  
 QY 2896 tctccattagagctgaagaagaatgaccactgcttctgtataacagctgtgtctttag 2955  
 DB 418 TCTCCTNTTAGGCGCTCAAAAGAAATGACCTACGTTTCTGTATACANCTGTGTTCTTGTAT 359  
 QY 2956 gtgtgttactgtacacagaagtggtgcaactgaagctctgtgtgtgtgtcgtatgaa 3015  
 DB 358 GTTGTGNTACTGNNACACAGAAAGTGTGTCNCGAGCGCTGTGCGTGGCCGTATGAAA 299  
 QY 3016 acctgtgagccctgtgagtgtaagctactgtcttccattcaattgtttacgctggaatttct 3075  
 DB 298 GCTGTGTAGCCCTGCGAGTTAAGTACTGCTTCATTCATTGTTTACGCTGGAATTTTCT 239  
 QY 3076 ccccatggaatgtaagtaaaacttaagtggtgtcatcaataaactgtaataactaa 3132  
 DB 238 CCCCATGGAATTAAGTAAACTTAAGTGTGTTGTCATCATATTAATGTTACTTAA 182  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 13  
 Z79981/c  
 ID Z79981 standard; cDNA; 660 BP.  
 XX  
 AC Z79981;  
 XX  
 DT 07-APR-2000 (first entry)  
 XX  
 DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:65.  
 XX  
 KW Human; gene expression product; diagnosis; tumour; colon cancer;  
 KW colorectal adenocarcinoma; cell line SW480; cell proliferation;  
 KW cytosstatic; sarcoma; breast cancer; neoplasia; dysplasia;  
 KW hyperplasia; ds.  
 KW  
 OS Homo sapiens.  
 XX  
 XX M09964576-A2.  
 XX  
 XX  
 XX 16-DEC-1999.  
 XX  
 PF 09-JUN-1999; 99WO-1B01062.  
 XX  
 PR 10-JUN-1998; 98US-0088801.  
 XX  
 PA (FARB ) BAYER CORP.  
 XX  
 PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;  
 PI Carroll E, Catino TJ, Dertl A, Ford DM, Lewis ME, Monahan JE;  
 PI Schlegel R;  
 XX  
 DR WPI: 2000-087220/07.  
 XX  
 XX Novel nucleic acids, used to develop products for the diagnosis and  
 PT treatment of disorders involving unwanted cell proliferation,  
 PT particularly cancers, especially colon cancer -  
 PS  
 PS Claim 1: Page 158; 469pp: English.  
 XX  
 CC Z79917 to Z80766 represent double stranded cDNA clones isolated from the  
 CC human colorectal adenocarcinoma (colon cancer) cell line SW480. The  
 CC cDNA clones can be used to generate antisense oligonucleotides which  
 CC can be used for antisense therapy. Methods and products from the present  
 CC invention can be used for identifying and/or classifying cancerous cells  
 CC present in a human tumour, particularly in solid tumours, e.g. carcinomas  
 CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used  
 CC for developing agents for the diagnosis and treatment of disorders  
 CC involving unwanted cell proliferation, such as neoplasia, dysplasia or  
 CC hyperplasia.  
 CC  
 CC Sequence 660 BP: 133 A; 172 C; 159 G; 174 T; 22 other;  
 SQ

Query Match 13.8%; Score 433.8; DB 21; Length 660;





---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 4, 2001, 16:38:46 ; Search time 313.08 seconds  
(Without alignments)  
1617.368 Million cell updates/sec

Title: US-09-316-048-1

Perfect score: 3142  
Sequence: 1 actctctcgcggtcgcggtg.....taatactataaaaaaaaaa 3142

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata/2/ina/6\_COMB.seq:\*  
4: /cgn2\_6/prodata/2/ina/PCrus\_COMB.seq:\*  
5: /cgn2\_6/prodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	70.8	2.3	7218	1	US-08-232-463-14 Sequence 14, Appl
2	68	2.2	7218	1	US-08-232-463-14 Sequence 14, Appl
3	52	1.7	12001	1	US-08-458-568A-11 Sequence 11, Appl
4	51.2	1.6	53526	3	US-08-658-136-2 Sequence 2, Appl
5	51.2	1.6	53577	3	US-08-658-136-1 Sequence 1, Appl
6	47.6	1.5	152331	3	US-09-128-155-16 Patent No. 5196516
7	45	1.4	4897	5	5196516-7 Patent No. 5196516
8	44.6	1.4	1558	1	US-08-455-550-7 Sequence 7, Appl
9	44.4	1.4	538	2	US-08-332-766A-24 Sequence 24, Appl
10	44.4	1.4	3918	3	US-08-836-329-1 Sequence 1, Appl
11	44.2	1.4	12001	1	US-08-458-568A-11 Sequence 11, Appl
12	41.4	1.3	31571	1	US-08-323-443B-1 Sequence 1, Appl
13	41	1.3	87350	3	US-08-781-891-79 Sequence 79, Appl
14	40.8	1.3	24979	2	US-08-147-777-3 Sequence 3, Appl
15	40.8	1.3	24979	3	US-08-452-872-3 Sequence 3, Appl
16	40.8	1.3	24979	4	PCT-US93-03985-3 Sequence 3, Appl
17	40.4	1.3	1707	1	US-08-790-309-1 Sequence 1, Appl
18	40.4	1.3	1710	2	US-08-912-794-1 Sequence 1, Appl
19	40.2	1.3	32207	2	US-08-770-379-20 Sequence 20, Appl
20	40	1.3	2721	5	5215881-2 Patent No. 5215881
21	40	1.3	8438	1	US-07-945-283-1 Sequence 1, Appl
22	39.6	1.3	1931	2	US-09-130-114-2 Sequence 2, Appl
23	39.6	1.3	2580	2	US-09-050-863-2 Sequence 2, Appl
24	39.6	1.3	2980	2	US-08-463-081B-13 Sequence 13, Appl
25	39.6	1.3	2980	2	US-08-461-379A-13 Sequence 13, Appl
26	39.6	1.3	2980	2	US-08-462-390B-13 Sequence 13, Appl
27	39.6	1.3	2980	2	US-08-463-074B-13 Sequence 13, Appl
28	39.6	1.3	2980	3	US-08-465-585C-13 Sequence 13, Appl

29	39.6	1.3	2980	3	US-08-652-446-13 Sequence 13, Appl
30	39.6	1.3	5452	2	US-09-130-114-1 Sequence 1, Appl
31	39.6	1.3	10596	1	US-07-884-811-15 Sequence 15, Appl
32	39.6	1.3	10596	1	US-07-885-971-15 Sequence 15, Appl
33	39.6	1.3	10596	1	US-08-087-783A-15 Sequence 15, Appl
34	39.6	1.3	10596	1	US-08-194-088B-15 Sequence 15, Appl
35	39.6	1.3	10596	2	US-08-194-087-15 Sequence 15, Appl
36	39.6	1.3	10596	4	PCT-US93-04648-15 Sequence 15, Appl
37	39.4	1.3	287	2	US-08-332-766A-27 Sequence 27, Appl
38	39.4	1.3	246240	2	US-08-724-394A-20 Sequence 20, Appl
39	39.4	1.3	246240	2	US-08-724-394A-21 Sequence 21, Appl
40	39.4	1.3	246240	2	US-08-724-394A-22 Sequence 22, Appl
41	39.2	1.2	530	3	US-08-758-662-4 Sequence 4, Appl
42	39.2	1.2	4524	2	US-08-845-998-7 Sequence 7, Appl
43	39.2	1.2	4524	3	US-09-206-537-7 Sequence 7, Appl
44	39	1.2	1456	3	US-09-308-406-1 Sequence 1, Appl
45	39	1.2	1860	2	US-08-331-644-3 Sequence 3, Appl

## ALIGNMENTS

RESULT 1  
US-08-232-463-14  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-fls  
; US-08-232-463-14  
Query Match 2.3%; Score 70.8; DB 1; Length 7218;













```
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,872
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/147,777
FILING DATE:
APPLICATION NUMBER: 07/876,289
FILING DATE: April 30, 1992
APPLICATION NUMBER: Unassigned (204/144)
FILING DATE: October 29, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 204/153
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ. ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 24979 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-452-872-3
```

```
Query Match      1.3%; Score 40.8; DB 3; Length 24979;
Best Local Similarity 59.7%; Pred. No. 1.1;
Matches 86; Conservative 0; Mismatches 57; Indels 1; Gaps 1;

QY 54 tggcggtctctcgcgcgtccgtcttgcacgcctcccccccgccggcgagcgc 113
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6957 TgCTCCTCTGgCTCTGCTCGGCGCTCTCTCTCTCTCTCTCTCTCTCGCGC 7016

QY 114 ccgcctccccc-accggcactccgcctctcccttcgctcccttcctctccctt 172
    | ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7017 GCCTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCATC 7076

QY 173 ttctctctctctccctccctc 196
    || |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7077 TTCTCCTCTTTTCTCTCTCTCTCTC 7100
```

Search completed: April 5, 2001, 08:45:28  
Job time: 58002 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2001, 00:29:19 ; Search time 11371 Seconds  
(without alignments)  
1982.509 Million cell updates/sec

Title: US-09-316-048-3

Perfect score: 3217

Sequence: 1 actccgcgcggcgtgcggg.....taatactaaaaaaaaa 3217

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: em\_est4a:\*  
42: em\_estfun:\*  
43: em\_esthum1:\*

44: em\_esthum2:\*  
45: em\_esthum3:\*  
46: em\_esthum4:\*  
47: em\_esthum5:\*  
48: em\_esthum6:\*  
49: em\_esthum7:\*  
50: em\_esthum8:\*  
51: em\_esthum9:\*  
52: em\_esthum10:\*  
53: em\_esthum11:\*  
54: em\_esthum12:\*  
55: em\_esthum13:\*  
56: em\_esthum14:\*  
57: em\_esthum15:\*  
58: em\_esthum16:\*  
59: em\_esthum17:\*  
60: em\_esthum18:\*  
61: em\_esthum19:\*  
62: em\_esthum20:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estov1:\*  
68: em\_estov2:\*  
69: em\_estp11:\*  
70: em\_estp12:\*  
71: em\_estp13:\*  
72: em\_estp14:\*  
73: em\_estp15:\*  
74: em\_estro1:\*  
75: em\_estro2:\*  
76: em\_estro3:\*  
77: em\_estro4:\*  
78: em\_estro5:\*  
79: em\_estro6:\*  
80: em\_estro7:\*  
81: em\_estro8:\*  
82: em\_estro9:\*  
83: em\_estro10:\*  
84: em\_estro11:\*  
85: em\_estro12:\*  
86: em\_estro13:\*  
87: gb\_est41:\*  
88: gb\_est42:\*  
89: gb\_est43:\*  
90: gb\_est44:\*  
91: gb\_est45:\*  
92: gb\_est46:\*  
93: gb\_est47:\*  
94: gb\_est48:\*  
95: gb\_est49:\*  
96: gb\_est50:\*  
97: gb\_est51:\*  
98: gb\_est52:\*  
99: gb\_est53:\*  
100: gb\_est54:\*  
101: gb\_est55:\*  
102: gb\_est56:\*  
103: gb\_est57:\*  
104: gb\_est57:\*  
105: gb\_est58:\*  
106: gb\_est59:\*  
107: gb\_est60:\*  
108: gb\_est61:\*  
109: gb\_est62:\*  
110: gb\_est63:\*  
111: gb\_est64:\*  
112: em\_esthum21:\*  
113: em\_esthum22:\*  
114: em\_esthum23:\*  
115: em\_estom1:\*  
116: em\_estom2:\*

117: em\_estp16:\*  
 118: em\_estp17:\*  
 119: em\_estp18:\*  
 120: em\_estp19:\*  
 121: em\_estp20:\*  
 122: em\_estp21:\*  
 123: em\_estp22:\*  
 124: em\_estp23:\*  
 125: em\_estp24:\*  
 126: em\_estp25:\*  
 127: em\_estp26:\*  
 128: em\_estp27:\*  
 129: em\_estp28:\*  
 130: em\_estp29:\*  
 131: em\_estp30:\*  
 132: em\_estp31:\*  
 133: em\_estp32:\*  
 134: em\_estp33:\*  
 135: em\_estp34:\*  
 136: em\_estp35:\*  
 137: em\_estp36:\*  
 138: em\_estp37:\*  
 139: em\_estp38:\*  
 140: em\_estp39:\*  
 141: em\_estp40:\*  
 142: em\_estp41:\*  
 143: em\_estp42:\*  
 144: em\_estp43:\*  
 145: em\_estp44:\*  
 146: em\_estp45:\*  
 147: em\_estp46:\*  
 148: em\_estp47:\*  
 149: em\_estp48:\*  
 150: em\_estp49:\*  
 151: em\_estp50:\*  
 152: em\_estp51:\*  
 153: em\_estp52:\*  
 154: em\_estp53:\*  
 155: em\_estp54:\*  
 156: em\_estp55:\*  
 157: em\_estp56:\*  
 158: em\_estp57:\*  
 159: em\_estp58:\*  
 160: em\_estp59:\*  
 161: em\_estp60:\*  
 162: em\_estp61:\*  
 163: em\_estp62:\*  
 164: em\_estp63:\*  
 165: em\_estp64:\*  
 166: em\_estp65:\*  
 167: em\_estp66:\*  
 168: em\_estp67:\*  
 169: em\_estp68:\*  
 170: em\_estp69:\*  
 171: em\_estp70:\*  
 172: em\_estp71:\*  
 173: em\_estp72:\*  
 174: em\_estp73:\*  
 175: em\_estp74:\*  
 176: em\_estp75:\*  
 177: em\_estp76:\*  
 178: em\_estp77:\*  
 179: em\_estp78:\*  
 180: em\_estp79:\*  
 181: em\_estp80:\*  
 182: em\_estp81:\*  
 183: em\_estp82:\*  
 184: em\_estp83:\*  
 185: em\_estp84:\*  
 186: em\_estp85:\*  
 187: em\_estp86:\*  
 188: em\_estp87:\*  
 189: em\_estp88:\*

190: gb\_gss25:\*  
 191: gb\_gss26:\*  
 192: gb\_gss27:\*  
 193: gb\_gss28:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1757.8	54.6	1801	146	U69197 U69197 Soar
2	743.4	23.1	829	106	BE300155 600944602
3	701	21.8	1018	136	BE867390 601442470
4	684.8	21.3	831	106	BE299828 600944602
5	646.2	20.1	1036	137	BE889880 601512132
6	625.6	19.4	686	40	AM160378 au72c03.y
7	621.2	19.3	724	97	AM952360 EST964430
8	617.2	19.2	752	137	BE894295
9	606.4	18.8	821	107	BE379687
10	604	18.8	609	107	BE380123
11	603.8	18.7	674	38	AV707406
12	602.2	18.5	616	111	BE674157
13	594.4	18.5	594	40	AM162002
14	588.2	18.3	669	28	AL037151
15	576	17.9	612	38	AV695897
16	571.4	17.8	612	38	AV695897
17	565.4	17.6	749	110	BE617223
18	562.2	17.5	653	87	AM163731
19	555	17.3	609	136	BE810330
20	551.6	17.1	615	134	BE082712
21	550	17.1	550	107	BE348298
22	548	17.0	564	40	AM163206
23	542.4	16.9	584	136	BE813737
24	534	16.6	673	135	BE781223
25	532.2	16.5	573	3	AA206573
26	531.2	16.5	536	136	BE837505
27	531	16.5	866	136	BE868328
28	517.8	16.1	910	138	BE966164
29	513	15.9	543	134	BE089861
30	511.4	15.9	513	13	AA889669
31	510.4	15.8	512	27	AI983007
32	508.8	15.8	852	137	BE872137
33	500.6	15.6	673	109	BE541462
34	497.4	15.5	633	89	AM320695
35	488.2	15.2	569	3	AA191622
36	487	15.1	516	25	AI819766
37	483.4	15.0	485	25	AI802592
38	481	15.0	665	106	BE278193
39	480.2	14.9	872	135	BE779672
40	478.4	14.9	480	134	BE042598
41	475	14.8	494	89	AM351909
42	470.4	14.6	509	142	N31181
43	470	14.6	517	7	AA427366
44	470	14.6	517	92	AM580332
45	467.4	14.5	703	137	BE888420

## ALIGNMENTS

RESULT 1  
 LOCUS U69197 1801 bp mRNA  
 DEFINITION U69197 Scores infant brain INIB Homo sapiens cDNA clone 22368, mRNA  
 VERSION U69197.1 GI:2739420  
 KEYWORDS EST.  
 SOURCE human.



[illegible][illegible]



```

QY 244 agccagagtttaaacacccttaaccctcagaaactgaacaaagaacacattgttccgtg 303
    |||||||
DB 180 ACCGAGAGTTTATTAAACCTTAACCTCTCAGAACGAAACAAACATATGTTCTCTG 239
QY 304 aagccctctctttaaagaagaacataaaccctactgttagaactaaatgacatgtgca 363
    |||||||
DB 240 AACGCCCTCTTTTAAAAAAGAACATTAACCCCTACTAGAACATAATGACATGTGCA 299
QY 364 tgaaccttgaaaaaacaacatgtataagcctgttagaccctactactctgtagtcagtcga 423
    |||||||
DB 300 TGAACCTTGGAAAAAACCAATATTAACCTGTGACCTTACTCTCGATGACATGCA 359
QY 424 cctataactacaacatgagaggtgctctatcccccagagctacttlaacccattccag 483
    |||||||
DB 360 CCTATAACTACAAACATGAGAGAGGTGCTTATCCCGGAGGTACTTTTACCATTTCCAG 419
QY 484 ttccaccttacttatacaagtgtgaactctctgtggagagacagcaatttaattgcaag 543
    |||||||
DB 420 TTCCACCTTACTTATCAAGTGAACCTTCTGTGGAGACAGCAATTTAATGGCAAG 479
QY 544 gaaagaacaagacagctggaacacagatgctgctgccaacagctttagagatccctgca 603
    |||||||
DB 480 GAAAGCAAGACAGCTGCGAAACAGATGCTGTGCCAAACGTTGAGGATCTGCACGA 539
QY 604 atgagccctgcagagagctgaggtgtaagtgaagagaaatccgaagaagaataatctca 663
    |||||||
DB 540 ATGAGCCCTCTGCAGAGAGCTGAGGTGAATGGAAGAAATCCGAAAGAAAGAAATCTCA 599
QY 664 ataaatctgaataaagtcgaagtgttttagatgtgacctaacagcaactgtcgtgtaatt 723
    |||||||
DB 600 ATAAATCTGAATAATGAATCAAGTG- TTGAGATGACACTTAAACGAACTTGCTGTGAATT 658
QY 724 tgaaggtgagccggagaggtgagccacccacatagaagaacttgtgaccaggttctgg 783
    |||||||
DB 659 TCGAGGTGCGCCGGAGAGATGG-CCACCCACATGAGAACTTTGTGACCA--GGTTCGG 715
QY 784 ttggggaggttctgggggaaggtgaagggaaagaagaatccaagaagaataatgc 841
    |||||||
DB 716 GTGGGGAGT---AGGGGGACAGGTGACGGAAGCCAGAAAGTTCCAGAGAAATATGC 770

```

```

RESULT 5
LOCUS BE889880 1036 bp mRNA EST 29-SEP-2000
DEFINITION 601512132P1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3913590 5',
VERSION BE889880
KEYWORDS BE889880.1 GI:10347645
SOURCE EST.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 1036)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC
CNA Library Preparation: Life Technologies, Inc.
CNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.lnl.gov
Plate: LCM720 row: P column: 07
High quality sequence stop: 706.
Location/Qualifiers
1..1036
/organism="Homo sapiens"

```

```

/db_xref="taxon:9606"
/clone="IMAGE:3913590"
/clone_id="NIH_MGC_71"
/tissue_type="telomysarcoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: uterus; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."
BASE COUNT 296 a 230 c 220 g 260 t
ORIGIN

```

```

Query Match 20.1%; Score 646.2; DB 137; Length 1036;
Best Local Similarity 91.5%; Pred. No. 2.5e-148;
Matches 740; Conservative 0; Mismatches 63; Indels 6; Gaps 5;

```

```

QY 2290 cggcccccagagctaaagaagcagaagaataatgctcttactctgagcgagcccaac 2349
    |||||||
DB 1 CGGCCCCAGAGCTAAAGCACGAGAAATCAAAATGCTTCTACTCAGCGTACCACCAAC 60
QY 2350 ttctttagtgcagagcccccacacacctctgagtagaccacacatcacacagctt 2409
    |||||||
DB 61 TTTTCTAGGTGTCACAGGCCCCACACCTCTGCACTACCCACACATCCACATGCTTT 120
QY 2410 ctcttcaacagtgatctgtagtctttagtcttcaatttcttcttgatgataagacat 2469
    |||||||
DB 121 CTTTCCAAACAGTGAATCTGATTTAGTTTCAATTTCTTTTGGATTGATATGCACT 180
QY 2470 atataaaatttcaatttgaaatcttccaaattgtatcagtttaataagcaagtttgaa 2529
    |||||||
DB 181 ATATAAATTTTCATTTGAGATTTCTCAATGTATCTGTTAAATAGCACAGTTTGGAA 240
QY 2530 acttctgagagctgacttcaataatctaacagcaagaagatataatcagtgatg 2589
    |||||||
DB 241 ACTTGCTGAGACTACTTATTAATCAATATCTAACGACAAATATATATCTGTTATG 300
QY 2590 tggtagacatttatttcaattgactaaacccagagacagtttcaagtgatgcaaatgtgt 2649
    |||||||
DB 301 TGGTTAGACATTTTATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 360
QY 2650 gccctctggttcaagctgaacagctcctggaacttcaaaaacttgaaatgctccca 2709
    |||||||
DB 361 GCCCTCTGCTCAGCTGAAACAGTCTGACCTTTCAAAAACCTTAATTAATGATCTCCACA 420
QY 2710 gttgtataatgagcaatttgaagaatttgaacttgaatgataatgtgttcaattt 2769
    |||||||
DB 421 GTTGTATTAATTTGACAAATTTTGAATTTTAACTTTTGAATTTTGAATTTTGAATTTT 480
QY 2770 tattcaatttatttctgttaattgataacaaagacttaaatgaacttgaactcgttlt 2829
    |||||||
DB 481 TATTTCAATTTTATTTTGTATTAATGCAAAACAGACTTAATTAATTAATTAATTAATTA 540
QY 2830 aaagattttaaacaacatgtgatactatacatatgagctcttgagacttgaactc 2889
    |||||||
DB 541 AAAGATTTATTAATAAAACATTTGATGATATATATATGAGCTTTGAGACTTGAATTC 599
QY 2890 tacactaagatgatactctc- atgtagtcataataaactgcagagt- atttcca 2946
    |||||||
DB 600 TACACTTACAGATATGATCTCAATGATGATGATGATGATGATGATGATGATGATGATG 659
QY 2947 gagtgtcgaact- gtaataatactccatagagctggaagaagaatgaactagcttctc 3005
    |||||||
DB 660 AAGTGTGATATCTGTTTCTTACTTCTTCAATTAAGGAGGAAAGAAATGAGACTTATTCG 719
QY 3006 tgaatacagctgtgtgttcttgaatgtgtgttactgtacacagaagtgtg- tgcactga 3064
    |||||||
DB 720 TATAAAGTGTGTTGCTTGAGGTTGCGGGTTACTGCTGTAACAGAAAGTGTGTTCCCGCA 779
QY 3065 ggtctgcgtgtgtgtcgtatgaaacc 3093
    |||||||
DB 780 GGGTCTGATGTGGCGCGCTTGTATACC 808

```

FEATURES

source

```

RESULT      6
AM160378    686 bp      mRNA      EST      09-NOV-1999
LOCUS       au2c03.v1 Schneider fetal brain 00004 Homo sapiens cDNA clone
DEFINITION  IMAGE:2781796 5' similar to TR:095793 095793 STAU6EN PROTEIN. ;,
            mRNA sequence.
ACCESSION   AM160378
VERSION     AM160378.1 GI:6299411
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE   1 (bases 1 to 686)
AUTHORS     Hillier,L., Allen,M., Bowles,L., Dubuque,T., Gaisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            ,J., Moore,B., Schellenberg,K., Stepcoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
            WashU-NCI human EST Project
            Unpublished (1997)
            Other ESTs: au72c03.x1
            Contact: Wilson RK
            Washington University School of Medicine
            444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watsn.wustl.edu
            This clone is available royalty-free through LNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40RP from Glibco
            High quality sequence stop: 431.
            Location/Qualifiers
                1..686
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_image="2781796"
                /clone_lib="Schneider fetal brain 00004"
                /sex="male"
                /tissue.type="frontal lobe"
                /dev_stage="5 months post-conception"
                /lab_host="DH10B"
                /note="Organ: brain; Vector: pBluescript SK (Stratagene);
                Site_1: SstI; Site_2: XhoI; Double-stranded cDNA was
                prepared from human fetal brain tissue. 5' and 3'
                adaptors were used in cloning as follows: 5' adaptor
                sequence:
                5'-GAGAGAGAGAGAGAGCTCAAGATCTTAAATTAATATCCCCCCCCCC-3'
                and 3' adaptor sequence:
                5'-GAGAGAGAGAGAGAGCTCAAGATCTTAAATTAATATCCCCCCCCCC-3'
                size-selected for >0.5 kb inserts and has an average
                insert size estimated at 1.2 kb. This library was
                constructed using the CAP-trapper method for full-length
                enrichment and has not undergone amplification. Library
                was constructed by Dr. Claudio Schneider (LNCIB-Area
                Science Park, Trieste, Italy). "
BASE COUNT   165 a      213 c      150 g      157 t      1 others
ORIGIN
Query Match      19.4%; Score 625.6; DB 40; Length 686;
Best Local Similarity 97.2%; Pred. No. 2.7e-143;
Matches 668; Conservative 0; Mismatches 15; Indels 4; Gaps 3;
QY      18 gggagcgtgagc-gctctcagcgtttggc--ggcgcgtgcgctctctcgcgtcc 74
        |||||||
        1 gggcgccctgagcgccgttttcagcgtttggcgcggcgccgcctcctcctgcgtccc 60
        |||||||
QY      75 gctctcttgaccgctcccccgcgcgcgcgcgcgcgcgcgcctctccaaaggccactc 134
        |||||||
        61 gcttctttgaccgctc-cccccccgccggccggcgccgcctctctccacggccactc 119
        |||||||
QY      135 cgcctctccctccctctcgtccctctctctcctcctcttctctctctccctcc 194
        |||||||

```

```

Db      120 cgcctcttccctcccttcccttcccttcccttcccttcccttcccttcccttcc 179
QY      195 tcgcccgcacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 254
        |||||||
        180 tcgcccgcacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 239
        |||||||
QY      255 attaaccttaacctctcagaactgaacaagaacacactgtctctcgtgaacgcctct 314
        |||||||
        240 attaaccttaacctctcagaactgaacaagaacacactgtctctcgtgaacgcctct 299
        |||||||
QY      315 tttaaaagaagaagcaaacccctactgtagaactaaatgcaactgtgcaactgtga 374
        |||||||
        300 tttaaaagaagaagcaaacccctactgtagaactaaatgcaactgtgcaactgtga 359
        |||||||
QY      375 aaaaaaacatgataagcgtgttgacccttctcgcgatgcagccacctaactac 434
        |||||||
        360 aaaaaaacatgataagcgtgttgacccttctcgcgatgcagccacctaactac 419
        |||||||
QY      435 aacatgagagaggtgctatccccgcgaggtactttaccacatltccagttccacctta 494
        |||||||
        420 aacatgagagaggtgctatccccgcgaggtactttaccacatltccagttccacctta 479
        |||||||
QY      495 cttatcaagtgaacttctctgtggagagacagcaattaaigcagaagaagaaga 554
        |||||||
        480 cttatcaagtgaacttctctgtggagagacagcaattaaigcagaagaagaaga 539
        |||||||
QY      555 cagcgtcgaaacagcagatctgctgcgaacgcgttgagatccttgagatgagccctg 614
        |||||||
        540 cagcgtcgaaacagcagatctgctgcgaacgcgttgagatccttgagatgagccctg 599
        |||||||
QY      615 ccagaaagcctgaggtgaatgaagaagaatccgaagaagaagaatcctaataatctga 674
        |||||||
        600 ccagaaagcctgaggtgaatgaagaagaatccgaagaagaagaatcctaataatctga 659
        |||||||
QY      675 ataagtcaggtgtttgagattgcactt 701
        |||||||
        660 ataagtcaggtgtttgagattgcactt 686
        |||||||
RESULT      7
AM952360    724 bp      mRNA      EST      01-JUN-2000
LOCUS       EST364430 MAGE resequences, MAGB Homo sapiens cDNA, mRNA sequence.
DEFINITION  AM952360
ACCESSION   AM952360
VERSION     AM952360.1 GI:8142042
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
            1 (bases 1 to 724)
            Heyde,P., Qi,R., Abernathy,K., Dharrap,S., Gaspard,R., Gay,C., Holt
            ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
            Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
            unpublished (2000)
            Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnd@tigr.org
            Plates: 38
            Seq primer: Reverse.
            Location/Qualifiers
                1..724
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_lib="MAGE resequences, MAGB"
                /note="Vector: pBluescriptSkm"
BASE COUNT   198 a      179 c      189 g      158 t
ORIGIN

```







```

/clone_lib="NIH_MGC_53"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (TI phage-resistant)"
/Note="Organ: bladder; Vector: pDNR-LIB (Clontech);
Site:1: Still (ggccgctggcc); Site:2: Still (ggccatctggcc
); Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCCATTATGGC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCCGCGGCGGCGAGATG-dt(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.55 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
BASE COUNT      178 a      125 c      96 g      210 t
ORIGIN
Query Match      18.8%; Score 604; DB 107; Length 609;
Best Local Similarity 100.0%; Pred. No. 5.6e-138;
Matches 604; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2264 agccatgccagagagagcgttgcacggcccgagagctaaagcacagagaaataca 2323
Db 1 AGCCATGCCAGAGAGAGCGCTTGGACCGGCCCCAGAGCTAAAGACACAGAGAAATCA 60
OY 2324 atgctctctacacagcggagccaactttctagtgctgcagcgcccaaccctctgc 2383
Db 61 ATGCTTCTCTACACAGCGTACCCAACTTTCTAGTGTGCCAGCGGCCCAACCTCTGCG 120
OY 2384 agtacccacacacatcacccactgcttctctcccaacagtcgactgacttctatct 2443
Db 121 AGTACCCACACATCACCACGCTTCTCTTCCCAAGAGATCTGATTTCTAGTTTCA 180
OY 2444 tatttctcttgatgatgatgacactatataaaatttcaatttgagaattctcaatgt 2503
Db 181 TATTTTCTTTGATGATGATGACATATATAAATTTTCATTTGAGAAATTTCTCAAT 240
OY 2504 atctagttaaaagacaggttggaaactgtctgagactgacttataaataactaac 2563
Db 241 ATCTAGTTAAATGACACAGTTTGGAACTTGTCTGAGCTGACTTATCAATAATCTAAC 300
OY 2564 cgacaaagatcatatcatctgtatgtgttagaacatttcaattgacttaaccag 2623
Db 301 CCACAAAGATCATATCCATGTGTATGTGTGACATTTTATTCATGACTAACCCAG 360
OY 2624 gacagcttcagtgatgacaaattgtgtccctctgtgtcagctgaaacagctctga 2683
Db 361 GACAGCTTTCAGTGATGCAAAATGTGTGCTCTGTGCTGAGTGAACAACCTCTGAG 420
OY 2684 caaaacacctgaataagctctccacagctgtataaattggacaatttgaatttaaac 2743
Db 421 CAAAACCTTGATATAGTCTCCACAGTTGTATTAATTTGGCAATTTTAAAC 480
OY 2744 tttagatgatcatctgtgtccatttcaatttcaatttcttgaatgacaaacaga 2803
Db 481 TTTAGATGATCATTTTGGTTCATTTTATTTATTTTATTTTGTATGCAACAGAGA 540
OY 2804 cttaaagtaaccttgatctctgttttaagaattataaaacattgtgatactatac 2863
Db 541 CTTAATGAACTTGTATCTCTTTTAAAGATTAATAAACAATTTGTATCTATACAT 600
OY 2864 atgg 2867
Db 601 ATGG 604

```

```

RESULT 11
LOCUS BE613972 686 bp mRNA EST 24-AUG-2000
DEFINITION 6015040931 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3905783 5',
ACCESSION BE613972

```

```

VERSION BE613972.1 GI:9895569
EST.
SOURCE human.
ORGANISM Homo sapiens
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 686)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LCM700 row: j column: 24
High quality sequence stop: 675.
Location/Qualifiers
1. 686
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3905783"
/clonelib="NIH_MGC_71"
/tissue_type="telomysarcoma"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: uterus; Vector: pCMV-Sport6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."
BASE COUNT      226 a      156 c      196 g      108 t
ORIGIN
Query Match      18.8%; Score 603.8; DB 110; Length 686;
Best Local Similarity 97.1%; Pred. No. 6.4e-138;
Matches 669; Conservative 0; Mismatches 12; Indels 8; Gaps 5;
OY 565 aacacgatgctgctgcacaaagcgttgagatctctgcagaatgagccctgcagagagc 624
Db 1 AACAGATGCTGCTGCTCCAAAGCGTTGAGGATCTCAGATGAGCCCTGCCAGAGAGCG 60
OY 625 tggagtgatgagagagagatccgaagaagaataatcctaataatcgaataatcaag 684
Db 61 TGGAGGTGATGAGAGAGATCCGAAGAGAAATCTCAATAATCTGAATAATGACAAG 120
OY 685 tcttggagatgacttgaacggaacttgctgtaatttcgaagtgagccggagagatg 744
Db 121 TGTTTGAGATTTGACTTAAACGGAAGTCTGCTGTAATTTGAGAGTGCGCGGAGAGTG 180
OY 745 gccaccccccacatgaagaacttltgacaaagtttcggttgggagttgttggggagag 804
Db 181 GCCACCCCAATGGAAGAAC-TTGTGACCAAGTTTGGT--GGGAGTTGTGGGGGAAG 237
OY 805 gtgaagggaagaagcagaagatttcaagaagaataatgcccgaatagctgttctggagagc 864
Db 238 GTGAAGGGGAAAGCAAGAGATTTCAAGAAATAATGCGCCATAGCTGTTCTTGAGAGCC 297
OY 865 tgaagaagtlacgcccctctgctgagttgaaagcagtaagcctagaatcaaaagaaga 924
Db 298 TGAAGAAGTTACCGCCCTGCTGCGAGTTGAAGCAAGTAAGCTTGAATCAAAAGAA 357
OY 925 caaaacccatagtcagacacagacagccagaatattgcccagggatcaatccgata 984
Db 358 CAAAACCCATATGTAAGCCACAGACAGAGCCAGAAATATGCCA--GGGATCAATCCGATTA 416
OY 985 gccgactggccagatccagcagcgaagaaagagagagagcagaagatacacgctccta 1044
Db 417 GCCGACTGGCCAGATCCAGCAGGCAAAAAGAGAGAGAGAGAGATACAGCTCTCTCA 476

```





```

Db      294 CTGAGAGTAAAGAGAGAAATCCGAAGAAATAATCTCAATAATCTGAAATAAGTCAA 235
QY      684 gtgtttgagattgaccttaaacggaacttcctcgtgtgaattcgaagtgcccggaagat 743
Db      234 GTGTTTGAATTGCACTTAACGGAACCTTGCTGTAATTTTCAGAGTGGCCCGGAGAGT 175
QY      744 gggccacccccacatgaagaactttgtaccaaagttcgtgtggaagttgtggaagaa 803
Db      174 GGGCCACCCCATATGAAGAACTTTGTGACCAAGGTTTGGTTGGGAGATTGTGGGGAA 115
QY      804 ggtgaagggaagaagaagaatttcaagaanaatgcgcgaatagctgtcttcaggag 863
Db      114 GGGAAGGGGAAAGCAAGAAATTTCAAGAAATAATCCGCATAGCTGTTCTTGAGAG 55
QY      864 ctgaagaagtccgcctcgtcgtgaattgaacgaagtaagactgaatcaaa 917
Db      54 CTGAAGAGATTACCGCCCTGCTCGCAATTGAAGAGTAAGCTGAGAAAAA 1

```

```

RESULT 15
LOCUS   AL037151      669 bp      mRNA      EST      29-FEB-2000
DEFINITION DKFZP56400264.L1 564 (synonym: hdbp2) Homo sapiens cDNA clone
ACCESSION AL037151
VERSION   AL037151.3 GI:5928071
KEYWORDS  EST
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 669)
AUTHORS   Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and Wiemann
          S.
TITLE     EST (Duesterhoeft, et al.)
JOURNAL   Unpublished (1999)
COMMENT   On Jul 7, 1999 this sequence version replaced gi:5866472.
          Contact: Duesterhoeft A
          MIPS
          Am Klopfersplitz 18a D-82152 Martinsried, Germany
          This is the 5' sequence of the clone insert
          Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
          Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
          sequenced by Olagen (Hilden/Germany) within the cDNA sequencing
          consortium of the German Genome Project.
          No sl sequence available.
          This clone (DKFZp56400264) is available at the RZPD in Berlin.
          Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
          Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
          Location/Qualifiers
            1..669
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="DKFZp56400264"
              /clone_lib="564 (synonym: hdbp2)"
              /tissue_type="brain"
              /dev_stage="fetal"
              /lab_host="xl-2blue"
              /note="Vector: pAMP1; Site_1: NotI; Site_2: SalI"

```

```

BASE COUNT      163 a      212 c      143 g      147 t      4 others
ORIGIN

```

```

Query Match      17.9%; Score 576; DB 28; Length 669;
Best Local Similarity 93.2%; Pred. No. 4.5e-131;
Matches 636; Conservative 0; Mismatches 24; Indels 8; Gaps 4;

```

```

QY      11 gggctcggcgagcgtgagc-gctctcaagcgtttgcgc--ggcgagtcgcgctctctc 67
Db      2  gggctcggcgagcgttgcgcgctcttgcgcgtttgcgc--ggcgagtcgcgctctctc 61
QY      68 ggtcccgctctcttgcgcgctctcccccggcgagcgcgagccgctctctccag 127

```

```

Db      62 GGGTCCCGCTTCCTTGACCGCCTCCCGCCCGGCGGCGGCGGCGGCTTCACAG 121
QY      128 ggcactcgcgctctccctccctcctcctcctcctcctcctcctcctcctcctcctcct 187
Db      122 GGCACCTCCGCTTTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 181
QY      188 cccctcctcgcgcgcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 247
Db      182 CCCCTCTCGCGCCACACGCGCCAGGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 241
QY      248 agagtttattaccacttaacctcctcagaactgaacaagaacacatgtctctgaaag 307
Db      242 AGAGTTTATTAAACACTTAACCTCAGAACTGAACAAAGCAACATTTGTTCTGGAAAG 301
QY      308 cccctcttttaaaaaagaagaagcctaaccctactgtagaactaaatgaactgtgcatgaa 367
Db      302 CCTCTTTTAAAGAAAGAAAGCAATTAACCTTACTGTGAATGAATGACACTGTGCAATGAA 361
QY      368 acttgaaaaaaaccaatgataagcctgttgacccttactctcgaatgcagtcaccta 427
Db      362 ACTTGGAAAAAACCAATGATATAAGCCTGTGACCTTACTCTCGATGCAATGCAACAA 421
QY      428 taactacaacatgaagaggtgcttatcccccgaagttacttaccacttccagttcc 487
Db      422 TAACCTAACACATGAGAGAGAGGTGCTTATCCCGAGAGGTACTTTTACCATTTCCAGTTCC 481
QY      488 accttacttatacgaagtgaacttctctgtggaagcagaacttaatgcaagaagaa 547
Db      482 ACCTTACTTATTAACAGTGAAGCTTCTGTGGAGGACACCAATTAATGCAAAAGAAA 541
QY      548 gacaagacagcgtcgaaacacagatgctgctcgaagcgttgaagatccctgcagaaatga 607
Db      542 GACAAGACAGGCTGCGAAACACGATGCTGCTGCCAAAGCGTTGAGATTCCTGCAAAATGA 601
QY      608 gccctcgtccagagagctgg--agtgaaatggaagaatcc--gaagaagaatctc 662
Db      602 ANCCCTGCCANMAAGGCTGGAGAGTGAATGAAGAAANAGATCCGAGAGAAANMAATCTC 661
QY      663 aataaatc 670
Db      662 AATTAATC 669

```

Search completed: April 5, 2001, 00:29:42  
 Job time: 53078 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2001, 08:45:28 ; Search time 313.08 Seconds

(without alignments)  
1655.975 Million cell updates/sec

Title: US-09-316-048-3

Perfect score: 3217  
Sequence: 1 actctctgccggcgctgcggg.....taatactaaaaaaaaa 3217

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/PCrUS.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	73	2.3	7218	1	US-08-332-463-14
2	68	2.1	7218	1	US-08-232-463-14
3	52	1.6	12001	1	US-08-458-568A-11
4	51.2	1.6	53526	3	US-08-658-136-2
5	51.2	1.6	53577	3	US-08-658-136-1
6	47.4	1.5	152331	3	US-09-128-155-16
7	45	1.4	4897	5	5196516-7
8	44.6	1.4	1558	1	US-08-455-550-7
9	44.4	1.4	538	2	US-08-332-766A-24
10	44.4	1.4	3918	3	US-08-836-329-1
11	44.2	1.4	12001	1	US-08-458-568A-11
12	41.4	1.3	31571	1	US-08-323-445B-11
13	41	1.3	87350	3	US-08-781-891-79
14	40.8	1.3	24979	2	US-08-147-777-3
15	40.8	1.3	24979	3	US-08-452-872-3
16	40.8	1.3	24979	4	PCr-US93-03985-3
17	40.4	1.3	1707	1	US-08-790-309-1
18	40.4	1.3	1710	2	US-08-912-794-1
19	40.2	1.2	32207	2	US-08-770-379-20
20	40	1.2	2721	5	5215881-2
21	40	1.2	8438	1	US-07-945-283-1
22	39.6	1.2	1931	2	US-09-130-114-2
23	39.6	1.2	2580	2	US-09-050-863-2
24	39.6	1.2	2980	2	US-08-463-081B-13
25	39.6	1.2	2980	2	US-08-461-379A-13
26	39.6	1.2	2980	2	US-08-462-390B-13
27	39.6	1.2	2980	2	US-08-463-074B-13
28	39.6	1.2	2980	3	US-08-465-585C-13

29	39.6	1.2	2980	3	US-08-652-446-13	Sequence 13, Appl
30	39.6	1.2	3452	2	US-09-130-114-1	Sequence 1, Appl
31	39.6	1.2	10596	1	US-07-884-811-15	Sequence 15, Appl
32	39.6	1.2	10596	1	US-07-885-971-15	Sequence 15, Appl
33	39.6	1.2	10596	1	US-08-087-783A-15	Sequence 15, Appl
34	39.6	1.2	10596	1	US-08-194-088B-15	Sequence 15, Appl
35	39.6	1.2	10596	2	US-08-194-087-15	Sequence 15, Appl
36	39.6	1.2	10596	4	PCr-US93-04648-15	Sequence 15, Appl
37	39.4	1.2	287	2	US-08-332-766A-27	Sequence 27, Appl
38	39.4	1.2	246240	2	US-08-724-394A-20	Sequence 20, Appl
39	39.4	1.2	246240	2	US-08-724-394A-21	Sequence 21, Appl
40	39.4	1.2	246240	2	US-08-724-394A-22	Sequence 22, Appl
41	39.2	1.2	530	3	US-08-758-662-4	Sequence 4, Appl
42	39.2	1.2	4524	2	US-08-845-998-7	Sequence 7, Appl
43	39.2	1.2	4524	3	US-09-206-537-7	Sequence 7, Appl
44	39	1.2	1456	3	US-09-308-406-1	Sequence 1, Appl
45	39	1.2	1860	2	US-08-331-644-3	Sequence 3, Appl

# ALIGNMENTS

RESULT 1  
US-08-232-463-14  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZ9c-Fls  
US-08-232-463-14

Query Match 2.3%, Score 73, DB 1, Length 7218;

















Query Match 2.28; Score 64.2; DB 1; Length 7218;





LENGTH: 53526 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-658-136-2

Query Match 1.7%; Score 49.6; DB 3; Length 53526;  
Best Local Similarity 53.9%; Pred. No. 0.00074;  
Matches 123; Conservative 0; Mismatches 104; Indels 1; Gaps 1;

QY 12 cctcccccgtacattctcgtcagagctgagcgagccagcgtcttcagcgtttgcgtg 71

Db 34527 cctccccctccctctcctccctccctccctccctccctccctccctccctcc 34586

QY 72 gctgctgcgtcgtctgtgtgctccctcttcttgagccccgcgtgagcgccgc 131

Db 34587 cctccccctccctccctccctccctccctccctccctccctccctccctcc 34646

QY 132 cttgcgtccgcacgtcgcgtcttcctcctcctgtgagccctttt-tcctgcgcgc 190

Db 34647 cttccccctcctcctcctccctccctccctccctccctccctccctccctcc 34706

QY 191 ttaactgtcttcttcaactctcgcgcgcgcgcgaagaccgcgcgcgc 238

Db 34707 cccccctccctccctccctccctccctccctccctccctccctccctcc 34754

RESULT 4  
US-08-658-136-1  
Sequence 1, Application US/08658136  
Patent No. 60711717

GENERAL INFORMATION:

APPLICANT: KLINGER, KATHERINE W

APPLICANT: LANDES, GREGORY M

APPLICANT: BURN, TIMOTHY C

APPLICANT: CONNORS, TIMOTHY D

APPLICANT: DACKOWSKI, WILLIAM

APPLICANT: GERMINO, GREGORY

APPLICANT: QIAN, FENG

TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE

NUMBER OF SEQUENCES: 58

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENZYME CORPORATION

STREET: ONE MOUNTAIN ROAD

CITY: FRAMINGHAM

STATE: MASSACHUSETTS

COUNTRY: USA

ZIP: 01701

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/658,136

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: LASSEN, ELIZABETH

REGISTRATION NUMBER: 31,845

REFERENCE/DOCKET NUMBER: GEN4-17.8

TELECOMMUNICATION INFORMATION:

TELEPHONE: 508-872-5415

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 53577 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-658-136-1

Query Match 1.7%; Score 49.6; DB 3; Length 53577;  
Best Local Similarity 53.9%; Pred. No. 0.00074;  
Matches 123; Conservative 0; Mismatches 104; Indels 1; Gaps 1;

QY 12 cctcccccgtacattctcgtcagagctgagcgagccagcgtcttcagcgtttgcgtg 71

Db 34522 cctccccctccctctcctccctccctccctccctccctccctccctccctcc 34581

QY 72 gctgctgcgtcgtctgtgtgctccctcttcttgagccccgcgtgagcgccgc 131

Db 34582 cctccccctccctccctccctccctccctccctccctccctccctccctcc 34641

QY 132 cttgcgtccgcacgtcgcgtcttcctcctcctgtgagccctttt-tcctgcgcgc 190

Db 34642 cttccccctcctcctcctccctccctccctccctccctccctccctccctcc 34701

QY 191 ttaactgtcttcttcaactctcgcgcgcgcgcgaagaccgcgcgcgc 238

Db 34702 cccccctccctccctccctccctccctccctccctccctccctccctcc 34749

RESULT 5  
US-08-836-329-1  
Sequence 1, Application US/08836329  
Patent No. 6090546

GENERAL INFORMATION:

APPLICANT: METHOD FOR THE DETECTION OF RAS ONCOGENES,

TITLE OF INVENTION: In Particular The K-Ras Oncogene

NUMBER OF SEQUENCES: 16

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/836,329

FILING DATE:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 3918 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: exon

LOCATION: 825..947

FEATURE:

NAME/KEY: exon

LOCATION: 1388..1567

FEATURE:

NAME/KEY: exon

LOCATION: 2118..2278

FEATURE:

NAME/KEY: exon

LOCATION: 3034..3158

FEATURE:

NAME/KEY: exon

LOCATION: 3459..3616

US-08-836-329-1

Query Match 1.6%; Score 45.8; DB 3; Length 3918;  
Best Local Similarity 50.7%; Pred. No. 0.0017;  
Matches 110; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 11 cctcccccgtacattctcgtcagagctgagcgagccagcgtcttcagcgtttgcgtg 70

```

Db      101 CCCGCGAAGGGGCCGCCAGACCCGGCTCTCCAGCCGGCCCGGGTCCGCACCTTAGAACCCC   160
Oy      71 ggcgtgctgcgcgtcgtgtgttgtagctccccccttctctgagccccggcgctggaggcgccg   130
Db      161 CCACAGCCACCCTTCCTCCTCCGCCCGCGCGCCCGCCCGCTCCTCCCCCGCGCGCGCGCGG    220
Oy      131 ccttcgacgccgcaccatctgcacctcttcctctctcgttgtgagcccttttcttcgcgcgtc    190
Db      221 CCCCCCTCTTTCTCCCGCGCGGCGGTGCGCTCCCTCTCTCCCTCTCTCCCAACACCGCC    280
Oy      191 ttcactgtcctctcacctccctccgcgcgcgcaccaaga   227
Db      281 CTCAGCCGCTCCCTCTCTGTCTACGCGCGCTGTGAAGAAGA   317

RESULT        6
US-08-861-464-1
; Sequence 1, Application US/08861464
; Patent No. 5874210
GENERAL INFORMATION:
APPLICANT: Guarente, Leonard P.
APPLICANT: Austrilaco Jr., Nicanor
APPLICANT: Kennedy, Brian
TITLE OF INVENTION: Genes Determining Cellular Senescence
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/861,464
FILING DATE: 22-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/396,001
FILING DATE: 28-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/09351
FILING DATE: 15-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/107,408
FILING DATE: 16-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
TELECOMMUNICATION INFORMATION:
TELEPHONE: 781-861-6240
TELEFAX: 781-861-9540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1946 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 322..1671
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 322..1671
US-08-861-464-1

```

```

Query Match      1.5%. Score 43.6; DB 2; Length 1946;
Best Local Similarity 54.3%; Fred. NO. 0.005;
Matches    88; Conservative    0; Mismatches   74; Indels    0; Gaps    0;

OY  42 gccccagcgcgtcttcacgclttcgctggctgcgtctgttgyccctccc 101
DB  733 GCCACCACCGCTACCTCCGGCGGCTGTATGGAAGTGTGATCTGTTCCTGCCAAC 792
OY  102 tcttcgagccccgcgcgtgcggcgccgcgccttcgcctccgaactcgctctcc 161
DB  793 ACTACTTTGTCTCCTCACACGCCGCCGCCCTACTACTCTGCACGCCGCCCTCTTCCTCC 852
OY  162 ctctgctgcgccctttctcctgcgcgtcttcactgctt 203
DB  853 TCTTCCTCTCTCCTCCTCTTCCTCTTCCTCTCTTCCTGATTCT 894

RESULT          7
US-08-396-001-1
; Sequence 1, Application US/08396001
; Patent No. 59196518
; GENERAL INFORMATION:
; APPLICANT: Guarante, Leonard P.
; APPLICANT: Austriaco Jr., Nicanor
; APPLICANT: Claus, James
; APPLICANT: Cole, Francesca
; APPLICANT: Kennedy, Brian
; TITLE OF INVENTION: Genes Determining Cellular Senescence In
; TITLE OF INVENTION: Yeast
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millitia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/396.001
; FILING DATE: 28-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-6408A2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1946 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 322..1671
; US-08-396-001-1

Query Match      1.5%; Score 43.6; DB 2; Length 1946;
Best Local Similarity 54.3%; Pred. NO. 0.005;
Matches    88; Conservative    0; Mismatches   74; Indels    0; Gaps    0;

OY  42 gccccagcgcgtcttcacgclttcgctggctgcgtctgttgyccctccc 101

```



```

: CURRENT FILING DATE: 1998-11-09
: NUMBER OF SEQ ID NOS: 348
: SOFTWARE: FASTSEQ For Windows Version 3.0
: SEQ ID NO 248
: LENGTH: 1308
: TYPE: DNA
: ORGANISM: Rat
: US-09-186-930-248

```

Query Match	1.5%;	Score 41.8;	DB 3;	Length 1308;
Best Local Similarity	51.5%;	Pred. No. 0.013;		
Matches 120;	Conservative 0;	Mismatches 112;	Indels 1;	Gaps 1;

[illegible]

RESULT 11  
US-09-188-930-244/c  
; Sequence 244, Application US/09188930A  
; Patent No. 6150503

```

: GENERAL INFORMATION:
: APPLICANT: Watson, James D.
: APPLICANT: Strachan, Lorna
: APPLICANT: Sleeman, Matthew
: APPLICANT: Onrust, Rene
: APPLICANT: Murlson, James Greg
: TITLE OF INVENTION: Compositions Isolated From Skin Cells
: TITLE OF INVENTION: and Methods For Their Use
: FILE REFERENCE: 11000.1011c1
: CURRENT APPLICATION NUMBER: US/09/188,930A
: CURRENT FILING DATE: 1998-11-09
: NUMBER OF SEQ. ID NOS: 348
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 244
: LENGTH: 1421
: TYPE: DNA
: ORGANISM: Mouse
: FEATURE:
: NAME/KEY: unsure
: LOCATION: (1370)...(1370)
: NAME/KEY: unsure
: LOCATION: (1395)...(1395)
: US-09-188-930-244

```

Query Match	1.58;	Score 41.8;	DB 3;	Length 1421;
Best Local Similarity	51.58;	Pred. No. 0.014;		
Matches 120;	Conservative 0;	Mismatches 112;	Indels 1;	Gaps 1;

QY	38	gcggagccccagagccgcgttccttaagaggtttgtagcgtgctgctgtagtcgtgtgtagctcc	97
Db	893	gcttgcttccgggaacccctttagcctctctcgtgaagcctctctcgtcgaagacattccggctt	834
QY	98	cccttctcttgagcccccgcgcctgagggagccc_gccttcgcctccgcacatccgcctctt	156
Db	833	ctctttgatgacccccacatgattggggcctttccggggcctctccctcggccattgaaacaccttc	774
QY	157	ccctcctctgtlgtggcccttttctcctgcgcgttcttaactgtcttcttaacctcttcgc	216

Db 773 CTTCTCCTTCTCTCTCTCTCTCTCTCTCGAGACCACCTTACCTCTTTTGTCTCTCCCACT 71

Qy 217 GCGGCGCAAGACCGCGCGGCCCGGAGCAGCTCTGGGAAAGCAAGCCAGAAAT 269

Db 713 GCGCTGGCAGGGGCTTACCTGAGGACCGGAGGTTTCGATACACCAACAATCT 661

RESULT 12  
US-09-050-863-2/c  
; Sequence 2, Application US/09050863  
; Patent No. 611A111

```

1  GENERAL INFORMATION:
2  APPLICANT: Lao, Ying
3  APPLICANT: Hiang, Betty
4  APPLICANT: Payan, Don
5  TITLE OF INVENTION: Mammalian Protein Interaction Cloning
6  TITLE OF INVENTION: System
7  NUMBER OF SEQUENCES: 5
8  CORRESPONDENCE ADDRESS:
9  ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
10 STREET: 4 Embarcadero Center, Suite 3400
11 CITY: San Francisco
12 STATE: CA
13 COUNTRY: USA
14 ZIP: 94111-4187
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE: Floppy disk
17 COMPUTER: IBM PC compatible
18 OPERATING SYSTEM: PC-DOS/MS-DOS
19 SOFTWARE: PatentIn Release #1.0, Version #1.30
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: US/09/050,863
22 FILING DATE: 30-MAR-1998
23 CLASSIFICATION:
24 ATTORNEY/AGENT INFORMATION:
25 NAME: Silva, Robin M.
26 REGISTRATION NUMBER: 38,304
27 REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
28 TELECOMMUNICATION INFORMATION:
29 TELEPHONE: (415) 781-1189
30 TELEFAX: (415) 949-8711
31 INFORMATION FOR SEQ. ID NO.: 2:
32 SEQUENCE CHARACTERISTICS:
33 LENGTH: 2580 base pairs
34 TYPE: nucleic acid
35 STRANDEDNESS: unknown
36 TOPOLOGY: unknown
37 MOLECULE TYPE: DNA
38 US-09-050-863-2

```

Query Match	1.4%;	Score 40.6;	DB 3;	Length 2580;
Best Local Similarity	54.1%;	Pred. No. 0.05;		
Matches 126;	Conservative	0;	Mismatches 104;	Indels 3;
				Gaps 2

[illegible]

```
RESULT 13
US-09-130-114-1
; Sequence 1, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Dama, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Strably Expressing Genes
; FILE REFERENCE: 0867/ID903051
; CURRENT FILING DATE: US/09/130,114
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 5452
; TYPE: DNA
; ORGANISM: VEBNA
US-09-130-114-1

Query Match          1.4%; Score 40.6; DB 2; Length 5452;
Best Local Similarity 54.1%; Pred. No. 0.084;
Matches 126; Conservative 0; Mismatches 104; Indels 3; Gaps 2;

QY 6 ggcggcccccgcgcgtacattctctgcagcgtggcgcccgagccgccttcagcgtt 65
DB 1844 gcccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1903
QY 66 ggcgtgagctgctgcgtgctgtgtgctccccccttctctgaagcccgcgctggcgc 125
DB 1904 cctccgctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1962
QY 126 gccgcgcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 185
DB 1963 cctctcctcgtcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2020
QY 186 ccgctctacattctctctacactctcctcgcgcgcgcgcgcgcgcgcgcgcgcgc 238
DB 2021 cctctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2073

RESULT 14
US-07-884-811-15/c
; Sequence 15, Application US/07884811
; Patent No. 5316921
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: palin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/884,811
; FILING DATE: 19920518
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055

US-09-130-114-1
```

```
REFERENCE/DOCKET NUMBER: 755.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: Single
; TOPOLOGY: linear
US-07-884-811-15

Query Match          1.4%; Score 40.6; DB 1; Length 10596;
Best Local Similarity 54.1%; Pred. No. 0.13;
Matches 126; Conservative 0; Mismatches 104; Indels 3; Gaps 2;

QY 6 ggcggcccccgcgcgtacattctctgcagcgtggcgcccgagccgccttcagcgtt 65
DB 2541 gcccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2482
QY 66 ggcgtgagctgctgcgtgctgtgtgctccccccttctctgaagcccgcgctggcgc 125
DB 2481 CCTCCTGCTCCTGCTCCTGCTCCTGCTCCTGCTCCTGCTCCTGCTCCTGCTCCTG 2423
QY 126 gccgcgcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 185
DB 2422 CCTCCTGCTCCTGCTCCTGCTCCTGCTCCTGCTCCTGCTCCTGCTCCTGCTCCTG 2365
QY 186 ccgctctacattctctctacactctcctcgcgcgcgcgcgcgcgcgcgcgcgcgc 238
DB 2364 CCTCCTGCTCCTGCTCCTGCTCCTGCTCCTGCTCCTGCTCCTGCTCCTGCTCCTG 2312

RESULT 15
US-07-885-971-15/c
; Sequence 15, Application US/07885971
; Patent No. 5328837
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: palin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/885,971
; FILING DATE: 19920518
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2001, 08:32:35 ; Search time 701.87 Seconds  
(without alignments)  
1529.155 Million cell updates/sec

Title: US-09-316-048-9  
Perfect score: 2857  
Sequence: 1 cggcgccgcgcctcccccgcg.....ccaataatgtaactactaa 2857

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues  
Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: N\_Geneseq\_36:\*  
2: /SID6/gcgdata/geneseq/geneseqn/NA1980.DAT:\*  
3: /SID6/gcgdata/geneseq/geneseqn/NA1981.DAT:\*  
4: /SID6/gcgdata/geneseq/geneseqn/NA1982.DAT:\*  
5: /SID6/gcgdata/geneseq/geneseqn/NA1983.DAT:\*  
6: /SID6/gcgdata/geneseq/geneseqn/NA1984.DAT:\*  
7: /SID6/gcgdata/geneseq/geneseqn/NA1985.DAT:\*  
8: /SID6/gcgdata/geneseq/geneseqn/NA1986.DAT:\*  
9: /SID6/gcgdata/geneseq/geneseqn/NA1987.DAT:\*  
10: /SID6/gcgdata/geneseq/geneseqn/NA1988.DAT:\*  
11: /SID6/gcgdata/geneseq/geneseqn/NA1989.DAT:\*  
12: /SID6/gcgdata/geneseq/geneseqn/NA1990.DAT:\*  
13: /SID6/gcgdata/geneseq/geneseqn/NA1991.DAT:\*  
14: /SID6/gcgdata/geneseq/geneseqn/NA1992.DAT:\*  
15: /SID6/gcgdata/geneseq/geneseqn/NA1993.DAT:\*  
16: /SID6/gcgdata/geneseq/geneseqn/NA1994.DAT:\*  
17: /SID6/gcgdata/geneseq/geneseqn/NA1995.DAT:\*  
18: /SID6/gcgdata/geneseq/geneseqn/NA1996.DAT:\*  
19: /SID6/gcgdata/geneseq/geneseqn/NA1997.DAT:\*  
20: /SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:\*  
21: /SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:\*  
22: /SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:\*

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2846	99.6	2859	293328	Mouse staufer cDNA
2	1354.6	47.4	3142	293324	Human staufer cDNA
3	1269.8	44.4	3217	293323	Human staufer cDNA
4	1256.6	44.0	3190	X90786	DNA encoding human
5	1256	44.0	3506	293322	Human staufer cDNA
6	1239	43.4	3260	293327	Human staufer cDNA
7	858	30.0	1127	X39991	Prostate cancer as
8	588.8	20.6	773	X39992	Prostate cancer as
9	575.8	20.2	946	X40002	Prostate cancer as
10	570.4	20.0	844	X40000	Prostate cancer as
11	360.2	12.6	660	279981	Human colon cancer
12	306.4	10.7	727	280269	Human colon cancer

13	60.6	2.1	300	20	217757
14	60.6	2.1	300	20	214226
15	58	2.0	769	20	217411
16	53	1.9	825	20	X40001
17	52	1.8	794	20	X39993
18	49.6	1.7	53526	19	T94101
19	49.6	1.7	53577	17	T18551
20	49.6	1.7	53577	19	T94108
21	49	1.7	10732	21	A10594
22	48.4	1.7	305	16	T20075
23	46	1.6	1000	21	A02484
24	45.8	1.6	3917	17	T31723
25	45	1.6	936	20	X56375
26	45	1.6	1337	20	T17263
27	43.8	1.5	1102	20	Z16923
28	43.6	1.5	1946	16	O82981
29	43.4	1.5	2188	20	Z77506
30	43.4	1.5	12001	16	O76213
31	42	1.5	6741	21	A10595
32	41.8	1.5	1308	21	Z61775
33	41.8	1.5	1421	21	Z61771
34	41.4	1.4	2883	21	A37029
35	41.2	1.4	1218	21	A02488
36	40.8	1.4	10732	21	A10594
37	40.6	1.4	799	19	V55831
38	40.6	1.4	1593	21	A02504
39	40.6	1.4	1925	20	X90924
40	40.6	1.4	5452	20	X90923
41	40.6	1.4	8705	20	Z23778
42	40.6	1.4	9600	19	Z21683
43	40.6	1.4	10380	20	Z22248
44	40.6	1.4	10596	14	O51731
45	40.6	1.4	10596	17	T40348

## ALIGNMENTS

Human gene express  
Human gene express  
Human gene express  
Prostate cancer as  
Prostate cancer as  
Human PKD1 gene.  
Human polycystic k  
Human PKD1 locus b  
Gene encoding a su  
Human gene signatu  
Human colon cancer  
Human colon cancer  
K-ras oncogene. H  
Human DNA-dependen  
Human gene express  
Human gene express  
UTR1 gene. Saccha  
Human ovalatin tumo  
HSV L/ST region.  
Gene encoding a su  
CDNA encoding rat  
CDNA encoding mus  
Human Prol113 (UNO  
Human colon cancer  
Gene encoding a su  
Nucleotide sequenc  
Human colon cancer  
Epstein Barr Virus  
Anti-sense strand  
Vector pShuttle DN  
Vector plasmid PCM  
Nucleotide sequenc  
Plasmid pcISBON f  
Plasmid pcISBON f

RESULT 1	293328	standard: cDNA: 2859 BP.
ID	293328	
AC	293328;	
XX	XX	
DT	04-JUL-2000	(first entry)
XX	XX	
DE	Mouse staufer cDNA.	
XX	XX	
KW	Staufen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;	
KW	HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;	
KW	acquired immune deficiency syndrome; protease; mouse; ss.	
XX	XX	
OS	Mus musculus.	
XX	XX	
FH	Key	Location/Qualifiers
FT	CDS	325..1788
FT		/*tag= a
XX	XX	/product= Staufen protein
XX	XX	
PN	CA2238656-A1.	
XX	XX	
PD	22-NOV-1999.	
XX	XX	
PF	22-MAY-1998;	98CA-2238656.
XX	XX	
PR	22-MAY-1998;	98CA-2238656.
XX	XX	
PA	(UYMO-) UNIV MONTREAL.	
XX	XX	
PI	Duchaine T, Wickham L, Cohen EA, Luo M, Desgroselliers L;	
XX	PI Moulard AJ;	
XX	XX	
DR	WPI: 2000-246924/22.	

DR P-PSDB: Y83024.  
XX Isolated mammalian or Caenorhabditis elegans staufen proteins useful  
PT for treating retroviral infections especially human immunodeficiency  
PT virus infections (HIV)  
XX  
XX Claim 4: Figure 1c: 96pp; English.  
XX  
CC Staufen is a RNA binding protein which interacts with double stranded  
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA  
CC transport and localization. Mammalian SP contain multiple ds  
CC RNA-binding domains. Chimeric proteins comprising portions of  
CC staufen with proteins with RNase or protease activity can be used to  
CC treat virus infections. The RNase or protease activity of the fusion  
CC protein prevents proper maturation of the virus. Mammalian staufen  
CC proteins seem to recognise double stranded RNA structure rather than  
CC any sequence specific position. 2 bacterially-expressed fusion  
CC proteins used in an RNA-binding assay (his/hsp (human SP) and  
CC MBP/msp (murine SP)) both proteins strongly bound double stranded  
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and  
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the  
CC treatment of RNA virus infections, especially those caused by  
CC retroviruses, in particular human immunodeficiency virus.  
XX  
SQ Sequence 2859 BP: 747 A; 750 C; 712 G; 650 T; 0 other:

Query Match 99.6%; Score 2846; DB 21: Length 2859;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2857: Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 cggagcgagcccccctccctcaactctctgcaagctcgagggcccgagccacttcag 60  
DB 2 cggagcgagcccccctccctcaactctctgcaagctcgagggcccgagccacttcag 61  
OY 61 cgttgcgtctgctgctgcgtctgctgctgctgctgctgctgctgctgctgctgctg 120  
DB 62 cgttgcgtctgctgctgcgtctgctgctgctgctgctgctgctgctgctgctgctg 121  
OY 121 gggcgccgcctctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 180  
DB 122 gggcgccgcctctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 181  
OY 181 cctgcgccttcaacttcttcaactctctgcccgcgcgcgcgcgcgcgcgcgcgcgcgcg 240  
DB 182 cctgcgccttcaacttcttcaactctctgcccgcgcgcgcgcgcgcgcgcgcgcgcgcg 241  
OY 241 gacgagccttgaggagagcagcagagaatataagcttcaacttgaagctcaatgacatgt 300  
DB 242 gacgagccttgaggagagcagcagagaatataagcttcaacttgaagctcaatgacatgt 301  
OY 301 gctgtgaacctggaagaacccaatgtataagccgtgagccctcaactctcgatgacgt 360  
DB 302 gctgtgaacctggaagaacccaatgtataagccgtgagccctcaactctcgatgacgt 361  
OY 361 ccaactacagcatagcatgctgctgaggtgctatccccccagataacttttcccatctc 420  
DB 362 ccaactacagcatagcatgctgctgaggtgctatccccccagataacttttcccatctc 421  
OY 421 cagtcaccacttacttactacccaagtgtgagcttcctggtggcgagcagagtttaatggga 480  
DB 422 cagtcaccacttacttactacccaagtgtgagcttcctggtggcgagcagagtttaatggga 481  
OY 481 aaggaagaatagagaccacccgtgaaacagatgccccctgcccgtgctgaggaactctgc 540  
DB 482 aaggaagaatagagaccacccgtgaaacagatgccccctgcccgtgctgaggaactctgc 541  
OY 541 aagagtgaaacccctgcagaaaggttgaggttaaatggaagaagaacagagaagaacc 600  
DB 542 aagagtgaaacccctgcagaaaggttgaggttaaatggaagaagaacagagaagaacc 601  
OY 601 tcaataaatacggaaataagccaagtgtttgaaatgcgtgaagcggaattgtccctgtga 660  
DB 602 tcaataaatacggaaataagccaagtgtttgaaatgcgtgaagcggaattgtccctgtga 661

DB 602 tcaataaatacggaaataagccaagtgtttgaaatgcgtgaagcggaattgtccctgtga 661  
OY 661 atttgaagtgcccgaggagatgcccacacacataaagaacttttgaccagaggtt 720  
DB 662 atttgaagtgcccgaggagatgcccacacacataaagaacttttgaccagaggtt 721  
OY 721 cagttggggaatttgtaagggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 780  
DB 722 cagttggggaatttgtaagggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 781  
OY 781 cggcagaggtctgttctggaagcagcttaagaggtctgcccctctctgtctgtgagcgag 840  
DB 782 cggcagaggtctgttctggaagcagcttaagaggtctgcccctctctgtctgtgagcgag 841  
OY 841 tgaagccagagatcaagagaagaatctcagccacttgcaagaacagcccgattatggcc 900  
DB 842 tgaagccagagatcaagagaagaatctcagccacttgcaagaacagcccgattatggcc 901  
OY 901 aagagatgaatcctatctatgtagacttgcacagatccagcaggaacaaaggaaggaagc 960  
DB 902 aagagatgaatcctatctatgtagacttgcacagatccagcaggaacaaaggaaggaagc 961  
OY 961 cagagatagctcctcttcaagaacagaggtcttccagctgcgaagaggttctgtatgcag 1020  
DB 962 cagagatagctcctcttcaagaacagaggtcttccagctgcgaagaggttctgtatgcag 1021  
OY 1021 taaagttgggcatcacacttcagaagaagtggtgtaccataaagaagtggtccaaagccta 1080  
DB 1022 taaagttgggcatcacacttcagaagaagtggtgtaccataaagaagtggtccaaagccta 1081  
OY 1081 atgctgcgtgaagaacatgctgtgagatcctcgtgggttcaaaagttcccagcgagcctgtca 1140  
DB 1082 atgctgcgtgaagaacatgctgtgagatcctcgtgggttcaaaagttcccagcgagcctgtca 1141  
OY 1141 agccagcactcaaatcagaagaagaagctccagctaaagaacacagagacggaagaagaag 1200  
DB 1142 agccagcactcaaatcagaagaagaagctccagctaaagaacacagagacggaagaagaag 1201  
OY 1201 taaagtttttgaaccttgaccttgaggatggaatgaagaactagtaacaagaagcagagagat 1260  
DB 1202 taaagtttttgaaccttgaccttgaggatggaatgaagaactagtaacaagaagcagagagat 1261  
OY 1261 tcaagatgcttaltatctttagcatalcagcagctgcagcctggaattctcccaatggtccgg 1320  
DB 1262 tcaagatgcttaltatctttagcatalcagcagctgcagcctggaattctcccaatggtccgg 1321  
OY 1321 aagttgcccagagctgtcggggttaagtcagaagacacacacacagaatttccacagggcag 1380  
DB 1322 aagttgcccagagctgtcggggttaagtcagaagacacacacacagaatttccacagggcag 1381  
OY 1381 ctccaaatccctgcccagaagcagaagtaactgtacatagatagccgaaggtgtgtgtagcggg 1440  
DB 1382 ctccaaatccctgcccagaagcagaagtaactgtacatagatagccgaaggtgtgtgtagcggg 1441  
OY 1441 gcaactcgcccaagcagagacatctttaaagagtaacatctctcagggccagcgtacccc 1500  
DB 1442 gcaactcgcccaagcagagacatctttaaagagtaacatctctcagggccagcgtacccc 1501  
OY 1501 atggaacctcgcaactagaccctctgagcaactgtatctactcttccagagcccagggattcc 1560  
DB 1502 atggaacctcgcaactagaccctctgagcaactgtatctactcttccagagcccagggattcc 1561  
OY 1561 aggttgaaatacaagaatttcccaagaacaacaagaagaggtgtatctctcaact 1620  
DB 1562 aggttgaaatacaagaatttcccaagaacaacaagaagaggtgtatctctcaact 1621  
OY 1621 gctcttcaagccgctctcgtaagtcataagcagcaagagatggtgagttcctgtcaat 1680  
DB 1622 gctcttcaagccgctctcgtaagtcataagcagcaagagatggtgagttcctgtcaat 1681  
OY 1681 atagtgctgcacttgaaacttttaaagctgctcgtcgtgagtttgaccacaaagacagacagaga 1740  
DB 1682 atagtgctgcacttgaaacttttaaagctgctcgtcgtgagtttgaccacaaagacagacagaga 1741



Qy 2820 aactcactctctgtgcaccaataatgtcataactaa 2657  
 |||||  
 Db 2822 aactcactctctgtgcaccaataatgtcataactaa 2659

Query Match	47.48;	Score 1354.6;	DB 21;	Length 3142;
Best Local Similarity	84.68;	Pred. No. 0;		
Matches 1597;	Conservative 0;	Mismatches 269;	Indels 21;	Gaps 6

QY	23	accctcgcgcagcgtctgcyggcccccagacgcgtctcttaagcgtttgcgtctgcgtctgcgc	82
Db	1	actctctgcgcggcgtctgcggccgcctctgaagcgtctcttaagcgtttgcgcgcgc	58
QY	83	gtctcgtgcygcctcccccctctctctag-----cccgccgtgcggcgcgcgc	132
Db	59	gtctctctcgcgtctccgcgtctctcttgacgcgcctcccccgcgcgcgcgcgcgcgc	118
QY	133	ctcgctccgcgcacatccgcctctctccctccctcgtcgcgcgcctcttctctgcgcgtct	192
Db	119	tcctccaaagcgcactccgcctctctccctctccctctgcctctctctctctctcttcttc	178
QY	193	caattgctctctcaactctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	252
Db	179	tctctctctctccctccctccctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	237
QY	253	ggaagcagc	312
Db	238	agc	297
QY	313	aaagaaacacaaatctataaagccgcgtgagacccctacctctgcgcgcgcgcgcgcgcgc	372
Db	298	gaaaaaaacacaaatctataaagcgttgagacccctacctctgcgcgcgcgcgcgcgcgc	357
QY	373	atgcgcgtcgtctgagcgtctgcctatcccccagatactttaccattccagctccacct	432
Db	358	aaacaaatgagc	417
QY	433	tactctccaaagcttgagcgtctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	492
Db	418	tactctccaaagcttgagcgtctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	477
QY	493	gaccacccgcgtataaagacgc	552
Db	478	gacacgc	537
QY	553	tcgcagaaagctctgagcgtataatggaagagcagagcagagaaacccctcaataactgcg	612
Db	538	tcgcagagagcgtctgagcgtataatggaagagcagagcagagaaacccctcaataactgcg	597
QY	613	aaataaagc	672
Db	598	aaataaagc	657
QY	673	ccgc	732
Db	658	ccgc	717
QY	733	ctgtagc	792
Db	718	ctgtgc	777
QY	793	cttcgc	852
Db	778	cttcctgc	837
QY	853	tcgaagaaagaaagctcagcccccactctga-----agacagcccccgcgcgcgcgcgcgcgc	906
Db	838	tcgaagaaagaaagctcagcccccactctga-----agacagcccccgcgcgcgcgcgcgcgc	897
QY	907	tgaaatcctatagtagaactctgcacagatccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	966
Db	898	tcgaatcctatagtagaactctgcacagatccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	957
QY	967	acatgc	1026
Db	958	acacgc	1017
QY	1027	cttgccgc	1086
Db	1018	cttgccgc	1077
QY	1087	cttggaacatgcgtgagatctcgtgggttcacaaagcttcccaagcgcgcgcgcgcgcgcgcgc	1146

[illegible]

FT CDS 363..1853  
 FT /\*tag= a  
 FT /product= Staufen  
 PN CA2238656-A1.  
 XX 22-NOV-1999.  
 PD 22-MAY-1998; 98CA-2238656.  
 PF 22-MAY-1998; 98CA-2238656.  
 PR 22-MAY-1998; 98CA-2238656.  
 PA (UTMO-) UNIV MONTREAL.  
 XX Duchaine T, Wickham L, Cohen EA, Luo M, Desrosellers L;  
 PI Moulard AJ;  
 DR WPI: 2000-246924/22.  
 DR P-PSDB: Y83108.  
 XX  
 PT Isolated mammalian or Caenorhabditis elegans staufen proteins useful  
 PT for treating retroviral infections especially human immunodeficiency  
 PT virus infections (HIV)  
 PS Claim 4; Figure 1a; 96pp; English.  
 CC Staufen is a RNA binding protein which interacts with double stranded  
 CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA  
 CC transport and localization. Mammalian SP contain multiple ds  
 CC RNA-binding domains. Chimeric proteins comprising portions of  
 CC staufen with proteins with RNase or protease activity can be used to  
 CC treat virus infections. The RNase or protease activity of the fusion  
 CC protein prevents proper maturation of the virus. Mammalian staufen  
 CC proteins seem to recognise double stranded RNA structure rather than  
 CC any sequence specific position. 2 bacterially-expressed fusion  
 CC proteins used in an RNA-binding assay (his/NSP (human SP) and  
 CC MBP/msp (murine SP) both proteins strongly bound double stranded  
 CC RNA. Both fusion proteins also directly bound labeled ds RNAs and  
 CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the  
 CC treatment of RNA virus infections, especially those caused by  
 CC retroviruses, in particular human immunodeficiency virus. This  
 CC sequence is an alternatively spliced sequence to the one given in  
 CC Z93322.  
 XX  
 XX  
 XX Sequence 3217 BP: 889 A; 796 C; 742 G; 790 T; 0 other;

Query Match 44.4%; Score 1269.8; DB 21; Length 3217;  
 Best Local Similarity 81.1%; Pred. No. 0;  
 Matches 1590; Conservative 0; Mismatches 277; Indels 94; Gaps 6;

QY 23 acttcctgcagagctgcggcgcccgagccgctcttcacgagcttgctgctgctgcgc 82  
 DB 1 acttcctgcggcgctgcggcgcccgagctcttcacgagcttgctgctgctgcgc 58  
 QY 83 gctctgtgctgc 132  
 DB 59 gctctctgc 118  
 QY 133 ttcgcctgc 192  
 DB 119 tctctcagcgcaactcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 178  
 QY 193 caactgtctctcaactcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 252  
 DB 179 tctctctcctccctcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 238  
 QY 253 ggaagcagcc-----262  
 DB 239 gaagcagcagagttatataacacttaacctctcagaactgaacaagaacaatgtt 298  
 QY 263 -----agaaaglatagcttaccatctgaactcaatgact 298

DB 299 ccttgaaagcgcctctttaaanaaagaagcatalaaccctactgtagaactaatgact 358  
 QY 299 gctgtgaaacttgaaanaaaccatgtataagcccgctgagccctcaactctcgatga 358  
 DB 359 gtgcataaacttgtaaaaaaaccatgtataagccctgttgacccttaactctcgatga 418  
 QY 359 gtccactcaagactatgtagcgttggaagtgccctccccagataactttaccact 418  
 DB 419 gtccactcaactaacaacatgagaggtgtctatcccccgagtaactttaccact 478  
 QY 419 tccagtcccaacttacttacttaccagttgagctctccgtggcgagacagcaattatag 478  
 DB 479 tccagttccacttacttacttacttaccagttgagctctgtggaagagcagcaattatag 538  
 QY 479 gaaggaagaagatgagaccctgtgaacacgactgcccctgcgcgtgagactct 538  
 DB 539 caaaggaagaagaagaagcagctgcggaacacgactgctgtgcacaaagcttgagact 598  
 QY 539 gcaagtgaaacccctgcggaagagtggaagtgaaagagaagcagaggaagaa 598  
 DB 599 gcaagtgagccctgcggaagagtggaagtgaaagagaagcagaggaagaa 658  
 QY 599 cctcaataatcgggaataagccagtggttgaaatgcgtggaagcgaatttgcct 658  
 DB 659 tctcaataatcgggaataagtggaatgttggaatgttggaatgttggaatgttggaat 718  
 QY 659 gaatttgagtggtgcccggagagtgccacacacatgaaagaacttgtgacagaggt 718  
 DB 719 gaatttcgagtggtgcccggagagtgccacacacatgaaagaacttgtgacagaggt 778  
 QY 719 ttcagtggtggaattgttggaaggaaggaaggaaggaaggaatctccaagaaga 778  
 DB 779 ttcggttggtggtgtgtgtggaagtggaaggaaggaaggaatctccaagaaga 838  
 QY 779 tgcgcgcagagctgtctgtggaagcgttaggaggtgcgcacccctccctgctgtgagcg 838  
 DB 839 tgcgcgcagagctgtctgtggaagcgttaggaggtgcgcacccctccctgctgtgagcg 898  
 QY 839 agtgaagcccaaatgaagaagaagtagccactgca-----agacagcccgga 892  
 DB 899 agtgaagcccaaatgaagaagaagtagccactgca-----agacagcccgga 958  
 QY 893 tlatgccaagagatgatacttacttactgtagactgtgacagatccagcaggaagaaaga 952  
 DB 959 atatgccaagagagatcacttacttactgtagactgtgacagatccagcaggaagaaaga 1018  
 QY 953 gaaggaagcagatgatacttacttactgtagactgtgacagatccagcaggaagaaaga 1012  
 DB 1019 gaaggaagcagatgatacttacttactgtagactgtgacagatccagcaggaagaaaga 1078  
 QY 1013 gatgacagtgataaggttggaatcacttgcagagaagagtggtgacacataagaagtggc 1072  
 DB 1079 gatgacagtgataaggttggaatcacttgcagagaagagtggtgacacataagaagtggc 1138  
 QY 1073 caagcgttaactgtctgtaaaacatgctgtagatccctggggttcaaaagttccccaagca 1132  
 DB 1139 caagcgttaactgtctgtaaaacatgctgtagatccctggggttcaaaagttccccaagca 1198  
 QY 1133 gctctcagaagcagacatcaaatcagaagaagactccagtaagaagaacagagaga 1192  
 DB 1199 gctctcagaagcagacatcaaatcagaagaagactccagtaagaagaacagagaga 1258  
 QY 1193 aaggaagaatgaagtttttgaactagccctggtggttgtaaaatggaactagtataaga 1252  
 DB 1259 aaggaagaatgaagtttttgaactagccctggtggttgtaaaatggaactagtataaga 1318  
 QY 1253 cgaagagttcagatgacttacttactgtagactgtgacagcgtgcagcgtggaattcccaat 1312  
 DB 1319 gataagagttcagatgacttacttactgtagactgtgacagcgtgcagcgtggaattcccaat 1378  
 QY 1313 ggtgcggaagtggtccagagctgtcggtgtagtaagaagacacacacaaagaattcac 1372  
 DB 1379 ggtgcggaagtggtccagagctgtcggtgtagtaagaagacacacacaaagaattcac 1438



```

Db 965 tagccagcagcccgccgacatccagcagcaaaaagagagagccagagtaacagctct 1024
Qy 977 tacagaacaggtctctccagcagaggtcttgatgcaggttaaggttggcatca 1036
Db 1025 cccagagcagagcctcccgccgcccagaggttctgtagcaggttgaaggttgaacca 1084
Qy 1037 cactgcagaagagatggtgtaccataaagaagtgcccaagcgttaatgctgtgagacat 1096
Db 1085 cactgcagaagagagcagcagcaacaagaagtgcccaagcgttaatgctgtgagacat 1144
Qy 1097 gctgtagatctcgggtgttcaagttcccaagcagcagcctgcccagcagcactcaatc 1156
Db 1145 gctgtagatccttgggttttcaagttcccaagcagcagcctgcccagcagcactcaatc 1204
Qy 1157 aagaagagagactccagtaaaagaaacagagagagagagagagagagagagagagagag 1216
Db 1205 agagagagagagagagagagagagagagagagagagagagagagagagagagagagag 1264
Qy 1217 tagccctggagatgaanaatggaactagtaacagagagagagagagagagagagagagag 1276
Db 1265 tagcctctggagatgaanaatggaactagtaacagagagagagagagagagagagagagag 1324
Qy 1277 tagccatgcagcagctgcccagctggaattctcccatgtgtccggaagtgcccaagctgt 1336
Db 1325 aagtcacagcagctgcccagctggaattctcccatgtgtccggaagtgcccaagctgt 1384
Qy 1337 cgggtgttagtcaagagacacacacacacacacacacacacacacacacacacacacac 1396
Db 1385 agagagtagtcaagagacacacacacacacacacacacacacacacacacacacacac 1444
Qy 1397 ggcacagagagagagagagagagagagagagagagagagagagagagagagagagagag 1456
Db 1445 ggcacagagagagagagagagagagagagagagagagagagagagagagagagagagag 1504
Qy 1457 cgaagacatttaagaagatgaatctctcagcagcagcagcagcagcagcagcagcagcag 1516
Db 1505 cgaagacatttaagaagatgaatctctcagcagcagcagcagcagcagcagcagcagcag 1564
Qy 1517 accctctggcagcagctactactcttccagagccagagagatccaggttgaatacaaga 1576
Db 1565 accctctggcagcagctactactcttccagagagcagagagatccaggttgaatacaaga 1624
Qy 1577 ttctcccaagaacaagaagagagagagagagagagagagagagagagagagagagagagag 1636
Db 1625 ctctcccaaaaaaacaagaagagagagagagagagagagagagagagagagagagagagag 1684
Qy 1637 tctcgtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag 1696
Db 1685 tctcgtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag 1744
Qy 1697 catcttaagctgctgtctgaggttgacaacagagacagagagagagagagagagagagag 1756
Db 1745 catcttaagctgctgtctgaggttgacaacagagagagagagagagagagagagagagag 1804
Qy 1757 tggacacagtttcagcgttgagagagagagagagagagagagagagagagagagagagag 1816
Db 1805 cggacacacagtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1864
Qy 1817 -cccaacataatacgaanaatacagagagagagagagagagagagagagagagagagagag 1875
Db 1865 tcccaacataatacgaanaatacga-aactgcttgaanaatttggaatttctgtatcc 1923
Qy 1876 tccagtgaggccagagaca 1892
Db 1924 tccagtgaggccagagaga 1940

```

RESULT 5  
 Z93322  
 ID Z93322 standard; cDNA: 3506 BP.  
 XX  
 AC Z93322;

```

XX 04-JUL-2000 (first entry)
DT
XX
XX Human staufen cDNA.
DE
XX
XX Staufen: RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW acquired immune deficiency syndrome; protease; human; ss.
OS
XX Homo sapiens.
FH
FH Key
FT CDS Location/Qualifiers
FT 409..2142
FT /*tag= a
FT /*product= Staufen
FT
CA2238656-A1.
XX
XX 22-NOV-1999.
PD
XX
XX 22-MAY-1998; 98CA-2238656.
PF
XX
XX 22-MAY-1998; 98CA-2238656.
PR
XX
XX (UYMO-) UNIV MONTREAL.
PA
PI Duchaine T, Wickham L, Cohen EA, Luo M, Desrosiers L;
PI Moulard AJ;
DR
DR WPI: 2000-246924/22.
DR P-PSDB: Y83023.
XX
XX Isolated mammalian or Caenorhabditis elegans staufen proteins useful
PT for treating retroviral infections especially human immunodeficiency
PT virus infections (HIV)
XX
XX Claim 4; Figure 1a; 96pp; English.
XX
XX Staufen is a RNA binding protein which interacts with double stranded
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA
CC transport and localization. Mammalian SP contain multiple ds
CC RNA-binding domains. Chimeric proteins comprising portions of
CC staufen with proteins with RNase or protease activity can be used to
CC treat virus infections. The RNase or protease activity of the fusion
CC protein prevents proper maturation of the virus. Mammalian staufen
CC proteins seem to recognise double stranded RNA structure rather than
CC any sequence specific position. 2 bacterially-expressed fusion
CC proteins used in an RNA-binding assay (his/hsp (human SP) and
CC MBP/MSP (murine SP)) both proteins strongly bound double stranded
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the
CC treatment of RNA virus infections, especially those caused by
CC retroviruses, in particular human immunodeficiency virus.
XX
XX Sequence 3506 BP; 965 A; 875 C; 790 G; 876 T; 0 other;

```

Query Match 44.0%; Score 1256; DB 21; Length 3506;  
 Best Local Similarity 86.7%; Pred. No. 0;  
 Matches 1420; Conservative 0; Mismatches 210; Indels 8; Gaps 3;

```

Qy 262 cagaagaatagatctacacattgagctcaatgacgtgtgtgaaactggaagaanaa 321
Db 611 cagaagaatagatctacacattgagctcaatgacgtgtgtgaaactggaagaanaa 670
Qy 322 caatgtataagccgtgagacccctacactcgtgagtgacgtccactaagatgagcagtc 381
Db 671 caatgtataagccgtgagacccctacactcgtgagtgacgtccactaagatgagcagtc 730
Qy 382 gtgagagtgctatccccccagatactttaccattccagtcacacacttactacc 441
Db 731 gagagagtgctatccccccagatactttaccattccagtcacacacttactacc 790

```



CC transport and localization. Mammalian SP contain multiple ds  
 CC RNA-binding domains. Chimeric proteins comprising portions of  
 CC Stafen with proteins with RNase or protease activity can be used to  
 CC treat virus infections. The RNase or protease activity of the fusion  
 CC protein prevents proper maturation of the virus. Mammalian stafen  
 CC proteins seem to recognize double stranded RNA structure rather than  
 CC any sequence specific position. 2 bacteriophage-expressed fusion  
 CC proteins used in an RNA-binding assay (his/hsp (human SP) and  
 CC MBP/msp (murine SP)) both proteins strongly bound double stranded  
 CC RNA. Both fusion proteins also directly bound labeled ds RNAs and  
 CC RNA/DNA hybrids. Stafen proteins may therefore be useful in the  
 CC treatment of RNA virus infections, especially those caused by  
 CC retroviruses, in particular human immunodeficiency virus. This  
 CC sequence is an alternatively spliced sequence to the one given in  
 CC 293322.

XX Sequence 3260 BP; 907 A; 795 C; 759 G; 799 T; 0 other;

Query Match 43.4%; Score 1239; DB 21; Length 3260;  
 Best Local Similarity 86.5%; Pred. No. 0;  
 Matches 1416; Conservative 0; Mismatches 210; Indels 11; Gaps 4;

QY 263 agaaagtagcttctacattgagctcaatgacatgctgtgtgaaactggaagaaacc 322  
 ||||| ||| | ||| | ||| ||||| ||||| ||||| ||| |||||  
 Db agaaagcataaccctactgtagaactaaatgacatgctgtgaaactggaagaaacc 428  
 QY 323 aatgtataagcccgtagacacctactctcgatgacgacgacacacacacacacac 382  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db aatgtataagcccgtagacacctactctcgatgacgacgacacacacacacacac 485  
 QY 383 tggagtgctacatcccccagatacttaccattccagctccacaccttaccacac 442  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 486 aggaagtgcttaccctcccgaggaacttaccattccagctccacaccttaccacac 545  
 QY 443 agttagagctctccgtaggcgacagacagtttaattggaaagaaagaaatgagaccac 502  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 546 agttagagctctccgtaggcgacagacagtttaattggaaagaaagaaatgagaccac 605  
 QY 503 gaaacagatgcccccgccgtagcgtgagacatcgcagagatgaaacccctgcacagaa 562  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 606 gaaacagatgcccccgccgtagcgtgagacatcgcagagatgaaacccctgcacagaa 665  
 QY 666 gctgagagatgagatgagaaagaaacagaaagaaacacacacacacacacacacac 725  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 726 agttagttagatgacacttaaacggaactgctgtgaaattcgaggtgcccggagag 785  
 QY 683 tggccac 742  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 786 tggccac 845  
 QY 743 aggaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaa 802  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 846 aggtgaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaa 905  
 QY 803 gcttagagagctgacacccctccctgctgtgagagcgatgaaacacacacacacacac 862  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 906 gctgaaagagctgacacccctccctgctgtgagagcgatgaaacacacacacacacac 965  
 QY 863 aagtcagccacactgca-----agacagcccccggattatggccacagagatgaa 916  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 966 aacaac 1025  
 QY 917 tagtaactgtcacagatccagcagcagcagcagcagcagcagcagcagcagcagcagc 976  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1026 tagcagactgcccagatccagcagcagcagcagcagcagcagcagcagcagcagc 1085  
 QY 977 tacagaacagagctctccacgctgcagagaggttctgtgacgtgaaaggttgagcaca 1036  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1086 cacagaacagagccctcccgccgacagagagttgtgatgcaagtgtaaggttgaaacca 1145  
 QY 1037 cactgtaagaagagtggtgtacccaataagaagtggtccaaagctgaatgctgtagaacat 1096  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1146 cactgtagaagaagagtggtgtacccaataagaagtggtgtacccaataagaagagat 1205  
 QY 1097 gctgagatcctgtgtgtacaaagttcccccagcagcagcagcagcagcagcagcagcagc 1156  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1206 gctgagatcctgtgtgtacaaagttcccccagcagcagcagcagcagcagcagcagcagc 1265  
 QY 1157 agaagaagagctccagtaagaagaacacagagagacagagaaagtaacgtttttgaacc 1216  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1266 agggaggaagacccacataagaagaacacagagagatggaagaagtaaccttttgaacc 1325  
 QY 1217 tagcctcgggagatgaagaatggaactagtaacagagacagagagttcagagttgcttacc 1276  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1326 tgcctcgggagatgaagaatggaactagtaacagagagatggaagttcagagttgcttacc 1385  
 QY 1277 tagcatalcagacgtgcacacgtggaattcccccagctgtgacgagagttgcccagagctgt 1336  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1386 aagtcatalcagacgtgcacacgtggaattcccccagctgtgacgagagttgcccagagctgt 1445  
 QY 1337 cgggtgtagtaagaagaac 1396  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1446 aggaagtagtaagaagaac 1505  
 QY 1397 ggcacaggtatgacatgtaac 1456  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1506 ggcacaggtatgacatgtaac 1565  
 QY 1457 cgaagacatttaagaagtaacatctctccagcagcagcagcagcagcagcagcagcagcag 1516  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1566 cgaagacatttaagaagtaacatctctccagcagcagcagcagcagcagcagcagcagcag 1625  
 QY 1517 accctctgagcaactgtactacatctccagagccacagagatcccaagttgatacaaga 1576  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1626 accctctgagcaactgtactacatctccagagccacagagatcccaagttgatacaaga 1685  
 QY 1577 ttctccaaagaacaaagaagagtgatctctcattacacatgctccctcagacggcc 1636  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1686 ctctccaaagaacaaagaagagtgatctctcattacacatgctccctcagacggcc 1745  
 QY 1637 tctcgtcagcatgacgtgcagcagagtgtagagtcctgtcagatataatgctgacacacac 1696  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1746 tctcgtcagcatgacgtgcagcagagtgtagagtcctgtcagatataatgctgacacacac 1805  
 QY 1697 catcttaagctgctgtcgtgagtgacacacacacacacacacacacacacacacacac 1756  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1806 catcttaagctgctgtcgtgagtgacacacacacacacacacacacacacacacacac 1865  
 QY 1757 tggacacagtttcaagcgtgctgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 1816  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1866 cggacacagtttcaagcgtgctgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 1925  
 QY 1817 -cccaacataatacctgaataactggaacacacacacacacacacacacacacacacacac 1875  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1926 tcccaacataatacctgaataactggaacacacacacacacacacacacacacacacacac 1984  
 QY 1876 tccagtgaggccaagaca 1892  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1985 tccagtgaggccaagaca 2001  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7  
 X3991  
 ID X3991 standard; DNA; 1127 BP.  
 XX  
 AC X3991;  
 XX  
 DT 02-UTL-1999 (first entry)  
 XX  
 DE Prostate cancer associated gene.  
 XX

KW Cancer associated antigen; diagnosis; research; treatment; human;  
 KM breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
 KM prostate cancer; ss.  
 XX Homo sapiens.  
 OS  
 XX MO9904265-A2.  
 PN  
 XX 28-JAN-1999.  
 PD  
 XX 15-JUL-1998; 98WO-US14679.  
 PE  
 XX 22-JUN-1998; 98US-0102322.  
 PR 17-JUL-1997; 97US-0896164.  
 PR 10-OCT-1997; 97US-0061599.  
 PR 10-OCT-1997; 97US-0061765.  
 PR 10-OCT-1997; 97US-0948705.  
 PR 11-OCT-1997; 97GB-0021697.  
 PA (LUDWIG) LUDWIG INST CANCER RES.  
 PI Chen Y, Gout I, Sure A, O'Hare M, Ohta Y, Old LJ;  
 PI Pfeunsschuh M, Sahin U, Scanlan MJ, Stockert E;  
 PI Tureci O;  
 XX MPI: 1999-132448/11.  
 DR  
 XX  
 PT New isolated cancer associated nucleic acids and polypeptides -  
 PT isolated using sera from cancer patients, used to develop products  
 PT for the diagnosis, monitoring or treatment of cancers  
 PS  
 XX  
 PS Claim 67: Page 625-626; 787pp; English.  
 CC The invention relates to a method for diagnosing a disorder characterised  
 CC by expression of a human cancer associated antigen precursor coded for by  
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
 CC biological sample isolated from a subject with an agent that specifically  
 CC binds to the NAM, an expression product or a fragment of an expression  
 CC product complexed with an HLA molecule; and (b) determining the  
 CC interaction between the agent and the NAM or the expression product as a  
 CC determination of the disorder. The products and methods can be used in  
 CC the diagnosis, monitoring, research, or treatment of conditions  
 CC characterised by the expression of various cancer associated antigens.  
 CC The invention provides nucleic acid sequences and encoded polypeptides  
 CC which are cancer associated antigen precursors expressed in human breast  
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
 CC lung cancer.  
 CC  
 CC Sequence 1127 BP; 349 A; 265 C; 307 G; 203 T; 3 other;  
 XX  
 SO  
 Query Match 30.0%; Score 858; DB 20; Length 1127;  
 Best Local Similarity 86.2%; Pred. No. 1.4e-246;  
 Matches 972; Conservative 0; Mismatches 148; Indels 7; Gaps 2;  
 QY 460 gcggacagcagcttcaatggaagaataagaccacccctggaacaacgagtcgccctg 519  
 DB 1 gggagacagcaatttaatggaagaagaagacagagctgcgaaacagatgctgctg 60  
 QY 520 cccgtgcgctggaagactctgcagatgaaacccctgcagaagaagcttgaggtlaaatgaa 579  
 DB 61 ccaagcgctgtagagatcctgcagaaatgagccctgcagagagcgctgagtgaaatgaa 120  
 QY 580 gagaagcagagaagaagaacacccataataatcggaataagcccaatggtttgaaattgcgc 639  
 DB 121 gagaatccgaagaagaataatcctaataatctgaaataagtaatglttggattgagcac 180  
 QY 640 tgaagcgaatttcgcttgaaatttgaggtgcccgcgagaagtgagccacacacatga 699  
 DB 181 ttaacgcgaacttgccttgaaatttcgaggtgcccgcgagaagtgagccacacacatga 240  
 QY 700 agaacttctgacacaggttcaagtgggaaattctgaagggaagaagaaggaaagca 759  
 ||||||||||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

DB 241 agaacttctgacacaggttctcggttgggagttcttgagggaagtgaaaggaaagca 300  
 QY 760 agaagatctccagaagaagatcgccgaaggcgtcttcttgagcagcttgaaggctgcac 819  
 DB 301 agaagattccaagaanaaatccgcataatgctgtcttcttgaggaagctgaaagattacgc 360  
 QY 820 cccctcctgctgtgagcgagtgaaagcccaatcaagaagaagaagtcagccacttgcga 879  
 DB 361 cccctgcctgcagttgaacgagtaagcctagatccaagaagaagaacacacacatagta 420  
 QY 880 -----agacagcccccgtattatgccaaggtatgaatcctcttcttgtagacttgcaga 933  
 DB 421 agccacaagacaagccagagatattggccaaggtatgaatcctcttctgagcagcccca 480  
 QY 934 tccagcagacaaagaagaagaagcagcaggtacatgctccttcacagaaggtcttc 993  
 DB 481 tccaagcagcaaaaaaagaagaagccagagtaacagctccttcacagaagcagagccttc 540  
 QY 994 cagctgcagaggttctgtatgcaggtlaaaggttgggcatcacactgcagaagagttgg 1053  
 DB 541 cggcgcgcagggagtttgatgacaggtgaaggttggaaccacactgcagaagagcag 600  
 QY 1054 gtaccaataagaagtgcccaagcgttaatgctgctgagacacatgcttgagatcctgggt 1113  
 DB 601 gcaccaacaagaagtgcccaagcgaatgcagccgagacacatgcttgagatcctggtt 660  
 QY 1114 tcaagttcccccagcgcagcctgcagcagcagcactcaatacaagaagaagactccag 1173  
 DB 661 tcaaaagtcgcagcgcagcagccacaaacccgcactcaagtcagaagaagacacca 720  
 QY 1174 taagaaaaccaaagacaggaagaaglaagcttcttgaacctgacctggagatgaa 1233  
 DB 721 taagaaaaccaaaggtatgaaagaagaagtaaccttcttgaacctgacctggagatgaa 780  
 QY 1234 atggaactagtaacaagaagcagaggtltaagatgcttacttgaacacagcagctgc 1293  
 DB 781 atgggactagtaataagaagatgattcaagatgcttacttgaacacagcagctgc 840  
 QY 1294 cagctggaaattcccatctgtgctgcggaaggtgcacagcgttcgggttgaatcaagac 1353  
 DB 841 ctgcctggaattctcccaagtgccgaaggtgcacagcgtgtaagatlaagtaagac 900  
 QY 1354 accacaacaaagattccacaaagcagcagctccaatcctctgcagaagcaacgtaactgcga 1413  
 DB 901 atcacacaacaaagatttaccagagcagctcgaatcctgcagaagcgaaggaactgcga 960  
 QY 1414 tgatagccgagagttgttctgtaaggggacactgcgccacag-ccgagacattttaag 1472  
 DB 961 tgatagccgagagttgttctgtaaggggacactgcgccacagcccgagacattttaag 1020  
 QY 1473 agtaacatctcttaagcagctaccccatgtaactgcgactagaaccctctagacaaactg 1532  
 DB 1021 aataacatctcttaagcagctaccccatgtaactgcgactagaaccctctnagcaactg 1080  
 QY 1533 tactaccttccagagccaggaattccaggttgaatalaacaagaatt 1579  
 DB 1081 gactactttnagagttcagaggtatnaggttgaataaccaagactt 1127  
 RESULT 8  
 X39992  
 ID X39992 standard; DNA; 773 BP.  
 XX  
 AC X39992;  
 DT 02-JUL-1999 (first entry)  
 XX  
 DE Prostate cancer associated gene.  
 XX  
 KW Cancer associated antigen; diagnosis; research; treatment; human;  
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
 KW prostate cancer; ss.  
 XX









```

|||||
Db 236 TTGAGATGCTTATCTAAGATCATCAGAGCTGCTGCTGGAATTTCTCCATGTGCCC 177
Qy 1320 gaagttgccagagctgtcgaggtagtcaagacacacacacaaagattaccaggga 1379
Db 176 GAGGTCCCTCCAGCTTACGAGTATGTCACAGATCATCACACAAAGATTTTACAGGCA 117
Qy 1380 gtccaaatctgcagaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 1439
Db 116 GCTCCGAATCTGTCAGAGGCGACGATGCTCCATGATGAGCCGAGAGTGTGATGG 57
Qy 1440 ggcacctgcgcacacagacagacacattlaaagaatacatctcttcagccagt 1495
Db 56 GGCACCTGCGCCACAGCGACAGACATTTTAAGATATACATCTCTCAGGCCAGCT 1

RESULT 12
ID 280269 standard; cDNA; 727 BP.
XX
AC 280269;
XX
DT 07-APR-2000 (first entry)
XX
DE Human colon cancer cell line SM480 cDNA clone SEQ ID NO:353.
XX
KW Human; gene expression product; diagnosis; tumour; colon cancer;
KW colorectal adenocarcinoma; cell line SM480; cell proliferation;
KW cytosolic; sarcoma; breast cancer; neoplasia; dysplasia;
KW hyperplasia; ds.
XX
OS Homo sapiens.
XX
PN WO9964576-A2.
XX
PD 16-DEC-1999.
XX
PF 09-JUN-1999; 99WO-1B01062.
XX
PR 10-JUN-1998; 98US-0088801.
XX
PA (FARB) BAYER CORP.
XX
PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catano TJ, Derti A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;
XX
DR WPI; 2000-087220/07.
XX
PT Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer.
XX
PS Claim 15; Page 273; 469pp; English.
XX
CC 279917 to 280766 represent double stranded cDNA clones isolated from the
CC human colorectal adenocarcinoma (colon cancer) cell line SM480. The
CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g. carcinomas
CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used
CC for developing agents for the diagnosis and treatment of disorders
CC involving unwanted cell proliferation, such as neoplasia, dysplasia or
CC hyperplasia.
XX
SQ Sequence 727 BP; 194 A; 153 C; 191 G; 149 T; 40 other;

```

Query Match 10.78; Score 306.4; DB 21; Length 727;  
 Best Local Similarity 83.8%; Pred. No. 1.6e-81;  
 Matches 357; Conservative 0; Mismatches 68; Indels 1; Gaps 1;

```

Qy 403 gatatttaccattcccaagtcaccattactctaccagattgagctccgtgagcg 462
Db 1 gtaatttaccattcccaagtcaccattactctaccagattgagctccgtgagcg 60
Qy 463 gacacagatttaatgaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 522
Db 61 gacacagatttaatgaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 120
Qy 523 gtgcgtgaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 582
Db 121 aagcgtgaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 180
Qy 583 aagcgtgaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 642
Db 181 aatccgaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 240
Qy 643 agcgaatttgctgtgaatttgaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggt 702
Db 241 aacggaacttgctgtgaatttgaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggt 300
Qy 703 acttgtgaccaggttcaattgaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggt 762
Db 301 acttgtgaccaggttcaattgaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggt 360
Qy 763 agatctccaagaagaatg-cggccagaggtgtctctgagcagcctgaagagctgccacc 821
Db 361 agatttcaaaaaaaagccgcagcatagctgntcttgaagagctgaagaagaatccgnc 420
Qy 822 ctccct 827
Db 421 ctgct 426

```

```

RESULT 13
ID 217757 standard; cDNA; 300 BP.
XX
AC 217757;
XX
DT 12-OCT-1999 (first entry)
XX
DE Human gene expression product cDNA sequence SEQ ID NO:5230.
XX
KW Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9938972-A2.
XX
PD 05-AUG-1999.
XX
PF 28-JAN-1999; 99WO-US01619.
XX
PR 03-APR-1998; 98US-0080666.
PR 28-JAN-1998; 98US-0072910.
PR 24-FEB-1998; 98US-0075954.
PR 31-MAR-1998; 98US-0080114.
PR 03-APR-1998; 98US-0080515.
XX
PA (CHIR) CHIRON CORP.
PA (HSE-) HSECO INC.
XX
PI Ctkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Inliss MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
DR WPI; 1999-494092/41.
XX
PT Novel human genes and their expression products which are

```













```

QY      258 aggtgagatttgccttgctcgtccaggctgaatgacagtgcgatcggactca 317
         ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 206042 AGATGGAGATTTCGCTTGTGTTGCCAGCGCTGAAGTGCAATGGCCGGAACTC GCGCTCACCG 205983

QY      318 caacctcacctcccagaatl--cagcgattttccaccattcagcctccagataagctgaaat 375
         ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 205982 CAACCTCATCCACCAGGTTCCAAGCAATTCCTGCCTCACGCTCCAGATAGCTGGGAT 205923

QY      376 tacagatttattaaccac 394
         ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 205922 TACAGCATGCATCACAC 205904


RESULT          3
US-08-724-394A-22/C
; Sequence 22, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Krommal, Gregory S.
; APPLICANT: Lauert, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: TOWNSEND AND TOWNSEND AND CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Flits, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.COMTIG"
US-08-724-394A-22


Query Match              2.6%; Score 88.2; DB 2; Length 246240;
Best Local Similarity    82.0%; Pred. No. 3e-12;
Matches 114; Conservative 0; Mismatches 23; Indels 2; Gaps 1.
```















Fri Apr 6 15:25:47 2001

us-09-316-048-7.rni

---

Page 9



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2001, 00:30:10 ; Search time 11371 Seconds

(without alignments)  
2063.239 Million cell updates/sec

Title: US-09-316-048-7

Perfect score: 3348

Sequence: 1 actctcctgcgcggctgcggg.....taatactaaaaa 3348

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*

2: gb\_est2:\*

3: gb\_est3:\*

4: gb\_est4:\*

5: gb\_est5:\*

6: gb\_est6:\*

7: gb\_est7:\*

8: gb\_est8:\*

9: gb\_est9:\*

10: gb\_est10:\*

11: gb\_est11:\*

12: gb\_est12:\*

13: gb\_est13:\*

14: gb\_est14:\*

15: gb\_est15:\*

16: gb\_est16:\*

17: gb\_est17:\*

18: gb\_est18:\*

19: gb\_est19:\*

20: gb\_est20:\*

21: gb\_est21:\*

22: gb\_est22:\*

23: gb\_est23:\*

24: gb\_est24:\*

25: gb\_est25:\*

26: gb\_est26:\*

27: gb\_est27:\*

28: gb\_est28:\*

29: gb\_est29:\*

30: gb\_est30:\*

31: gb\_est31:\*

32: gb\_est32:\*

33: gb\_est33:\*

34: gb\_est34:\*

35: gb\_est35:\*

36: gb\_est36:\*

37: gb\_est37:\*

38: gb\_est38:\*

39: gb\_est39:\*

40: gb\_est40:\*

41: em\_est41:\*

42: em\_est42:\*

43: em\_est43:\*

44: em\_est44:\*

45: em\_est45:\*

46: em\_est46:\*

47: em\_est47:\*

48: em\_est48:\*

49: em\_est49:\*

50: em\_est50:\*

51: em\_est51:\*

52: em\_est52:\*

53: em\_est53:\*

54: em\_est54:\*

55: em\_est55:\*

56: em\_est56:\*

57: em\_est57:\*

58: em\_est58:\*

59: em\_est59:\*

60: em\_est60:\*

61: em\_est61:\*

62: em\_est62:\*

63: em\_est63:\*

64: em\_est64:\*

65: em\_est65:\*

66: em\_est66:\*

67: em\_est67:\*

68: em\_est68:\*

69: em\_est69:\*

70: em\_est70:\*

71: em\_est71:\*

72: em\_est72:\*

73: em\_est73:\*

74: em\_est74:\*

75: em\_est75:\*

76: em\_est76:\*

77: em\_est77:\*

78: em\_est78:\*

79: em\_est79:\*

80: em\_est80:\*

81: em\_est81:\*

82: em\_est82:\*

83: em\_est83:\*

84: em\_est84:\*

85: em\_est85:\*

86: em\_est86:\*

87: em\_est87:\*

88: em\_est88:\*

89: em\_est89:\*

90: em\_est90:\*

91: em\_est91:\*

92: em\_est92:\*

93: em\_est93:\*

94: em\_est94:\*

95: em\_est95:\*

96: em\_est96:\*

97: em\_est97:\*

98: em\_est98:\*

99: em\_est99:\*

100: em\_est100:\*

101: em\_est101:\*

102: em\_est102:\*

103: em\_est103:\*

104: em\_est104:\*

105: em\_est105:\*

106: em\_est106:\*

107: em\_est107:\*

108: em\_est108:\*

109: em\_est109:\*

110: em\_est110:\*

111: em\_est111:\*

112: em\_est112:\*

113: em\_est113:\*

114: em\_est114:\*

115: em\_est115:\*

116: em\_est116:\*

```

117: em_estp16:*
118: em_estp17:*
119: em_estp18:*
120: em_estp14:*
121: em_estp15:*
122: em_estp16:*
123: em_estp17:*
124: em_estp18:*
125: em_estp19:*
126: gb_est58:*
127: gb_est59:*
128: gb_est60:*
129: gb_est61:*
130: gb_est62:*
131: gb_est63:*
132: gb_est64:*
133: gb_est65:*
134: gb_est66:*
135: gb_est67:*
136: gb_est68:*
137: gb_est69:*
138: gb_est70:*
139: gb_est71:*
140: gb_est72:*
141: gb_est73:*
142: gb_est74:*
143: gb_est75:*
144: gb_est76:*
145: gb_est77:*
146: gb_est78:*
147: gb_est79:*
148: gb_est80:*
149: gb_est81:*
150: gb_est82:*
151: gb_est83:*
152: gb_est84:*
153: gb_est85:*
154: gb_est86:*
155: gb_est87:*
156: gb_est88:*
157: gb_est89:*
158: gb_est90:*
159: gb_est91:*
160: gb_est92:*
161: gb_est93:*
162: gb_est94:*
163: gb_est95:*
164: gb_est96:*
165: gb_est97:*
166: gb_est98:*
167: gb_est99:*
168: gb_est100:*
169: gb_est101:*
170: gb_est102:*
171: gb_est103:*
172: gb_est104:*
173: gb_est105:*
174: gb_est106:*
175: gb_est107:*
176: gb_est108:*
177: gb_est109:*
178: gb_est110:*
179: gb_est111:*
180: gb_est112:*
181: gb_est113:*
182: gb_est114:*
183: gb_est115:*
184: gb_est116:*
185: gb_est117:*
186: gb_est118:*
187: gb_est119:*
188: gb_est120:*
189: gb_est121:*

```

190: gb\_gss25:\*  
 191: gb\_gss26:\*  
 192: gb\_gss27:\*  
 193: gb\_gss28:\*

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1757.8	52.5	1801	146	U69197 U69197 Soar
2	701	20.9	1018	136	BE667390 BE667390
3	646.2	19.3	1036	137	BE889880 BE889880
4	621.2	18.6	724	97	AM952360 AM952360
5	617.2	18.4	752	137	BE894295 BE894295
6	610.4	18.2	829	106	BE300155 BE300155
7	606.4	18.1	821	107	BE379687 BE379687
8	604	18.0	609	107	BE380123 BE380123
9	603.8	18.0	686	110	BE613972 BE613972
10	602.2	18.0	674	38	AV707406 AV707406
11	599.8	17.9	665	106	BE278193 BE278193
12	594.4	17.6	594	111	BE674157 BE674157
13	588.2	17.6	594	40	AM162002 AM162002
14	571.4	17.1	612	38	AV695897 AV695897
15	565.4	16.9	749	110	BE617223 BE617223
16	555	16.6	609	136	BE810330 BE810330
17	551.6	16.5	615	134	BE082712 BE082712
18	550	16.4	550	107	BE348298 BE348298
19	548	16.4	564	40	AM163206 AM163206
20	542.4	16.2	584	136	BE813737 BE813737
21	540.6	16.1	831	106	BE298828 BE298828
22	534	15.9	673	135	BE781223 BE781223
23	532.2	15.9	573	3	AA206573 AA206573
24	531.2	15.9	536	136	BE837505 BE837505
25	531	15.9	866	136	BE868328 BE868328
26	528.8	15.8	581	106	BE278342 BE278342
27	517.8	15.5	910	138	BE966164 BE966164
28	513	15.3	543	134	BE089861 BE089861
29	511.4	15.3	513	13	AA889669 AA889669
30	510.4	15.2	512	27	AI988307 AI988307
31	508.8	15.2	852	137	BE872137 BE872137
32	500.6	15.0	673	109	BE541462 BE541462
33	497.4	14.9	633	89	AM320695 AM320695
34	488.2	14.6	569	3	AA191622 AA191622
35	487	14.5	516	25	AI819766 AI819766
36	483.4	14.4	485	25	AI802592 AI802592
37	481.4	14.4	686	40	AM160378 AM160378
38	478.4	14.3	480	134	BE042598 BE042598
39	475	14.2	494	89	AM351909 AM351909
40	470.4	14.1	509	142	N31181 N31181
41	470	14.0	517	7	AA427366 AA427366
42	470	14.0	517	92	AA580332 AA580332
43	467.4	14.0	703	137	BE888420 BE888420
44	457	13.6	724	38	AV700494 AV700494
45	455.2	13.6	561	141	H16250 H16250

## ALIGNMENTS

```

RESULT 1
LOCUS U69197 1801 bp mRNA
DEFINITION U69197 Scores Infant brain INIB Homo sapiens CDNA clone 22368, mRNA
ACCESSION U69197
VERSION U69197.1 GI:2739420
KEYWORDS EST.
SOURCE human.

```

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 1801)  
AUTHORS Volorio,S., Simon,G., Repetto,M., Cucciardì,M., Banfi,S., Borsani,  
G., Ballabio,A. and Zollo,M.  
TITLE Sequencing analysis of forty-eight human image cDNA clones similar  
to drosophila mutant protein  
JOURNAL DNA Seq. 9 (5-6), 307-315 (1998)  
MEDLINE 99452386  
COMMENT Contact: Zollo, Massimo  
Telethon Institute of Genetics and Medicine  
Via Olgettina 58, Milan, MI 20132, Italy  
Email: zollo@tigem.it.

FEATURES  
source  
1..1801  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="19p13.1"  
/clone="72368"  
/clone\_lib="Scares infant brain INIB"  
/sex="female"  
/dev\_stage="73 days post natal"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: whole brain; Vector: Lambda BA; Site: 1: Not  
I; Site: 2: Hind III; 1st strand cDNA was primed with a Not  
I - oligo(dT) primer 15';  
AACTGGAAGAATTCGGCGGAGGAAATTTTATTTTATTTT 3';  
double-stranded cDNA was ligated to Hind III adaptors  
(pharmacia), digested with Not I and directionally cloned  
into the Not I and Hind III sites of the Lambda BA vector.  
Library went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 503 a 409 c 386 g 503 t  
ORIGIN

Query Match 52.5%; Score 1757.8; DB 146; Length 1801;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1770; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1577 gctccgaatcctcgcgaagcgaagcgaatgcatgcatgtagccgagagtgctgtagtggg 1636  
DB 5 GCTCCGAATCCTCGCAAGGCGCAGGTAACATGATGATGATGATGATGATGATGATGATG 64  
QY 1637 ggaactcgc 1696  
DB 65 GGCACCTCGCCACAGCGGACGACATTTTAAAGAAATTAACATCTCTTCAGGCCACGTAACC 124  
QY 1697 catggacctctcaagagacctctgagcaactggaactatcttccagagtcgaagagatc 1756  
DB 125 CATGGACCTCTCACGAGACCTCTGAGCAACTGAGCTATCTTTCCAGAGTCCAGGATTC 184  
QY 1757 caagttgaatataaagactctcccaaaaacaagaagaagaatctgtatctatcat 1816  
DB 185 CAGGTGAATATACAAAGACTTCCCAAAAACAACAAGAACCAATTTGATCTCTTATCAAT 244  
QY 1817 tgcctcttaagccacctctgcatcgcgatgcatgtagcgaagatgtgagatcctgcat 1876  
DB 245 TGCCTCCTCAGCCACCTCTGATCAGCATGATCGCAAGGATGAGGTCTCTGCAT 304  
QY 1877 gataatgctgcgtgaatcatctaaagtgtgtctgagtcgaagtcgaacaagaatgagag 1936  
DB 305 GATATGCTCGCTGGAACATCTTAAAGTCTGTCTGATTTGACCAACAAATGACAGAG 364  
QY 1937 atgcacaagaacaggaacagcaatgctgctgtgtgtagaggtgtgtagaactcttcgagc 1996  
DB 365 ATGCCAAGAACAGGAACGAGCAATGTGTGTGGGAGGTGCTGAACCTTTCTGGGC 424  
QY 1997 catgaacatataaataatcccaacatatatactgaaataactgaaactgtcttgaataat 2056  
DB 425 CATGAACCATTTATAAATCCCAACATATATATGAAAAATCTGAACCTGCTTGAATAAT 484

QY 2057 tgggaattctgataacctccagctggccgagagagaacagctggtgtaaaagatctgtagcaga 2116  
DB 485 TGGAAATTTTGATATACCTCCAGTGGGCCGAGAGACACGCTGGTGAAGAGATGTGGGACGA 544  
QY 2117 gcagggaagaacaagaacaagaagaagcgcgtgtgccc-gcgtgactgtctgtgggt 2175  
DB 545 GCAGGGAAGCAACAGAAACAAAGAGAGCGGCTGTGGCGGGCTGACGTCTGTGGGGT 604  
QY 2176 ttgttgtagtggcaccctcggtgacctgctgctccctcgaataagagcgtgcctgtggg 2235  
DB 605 TTGTTGTGATGGCCACTCGGTGACCTGGCGTCCCTAAGCAATAGCAGTGCCTGTGGGG 664  
QY 2236 aagaagggctgc 2295  
DB 665 AAGAAAGCTGCCCCAGCCACTGTTCTCCCGGACACAGCATCATCACCTGTGGGCA 724  
QY 2296 cctccgtgtgtgtgtcttcttcttccctgtgtgaaagaagaacgcgcgcgcctctc 2355  
DB 725 CTTCCGTTGTGGCTTTTCTTTTCCCTGTGTGAAGAAAGACGCGACACCCCTTCTC 784  
QY 2356 aagctgctcaactcaagacacatctggagcaaaccttgagcccatgacagagagagcct 2415  
DB 785 AAGCTGGCTCACTGACACATGGGACAAACCTTGACAGCCCTGCGCAGAGAGAGGCT 844  
QY 2416 ttgacggcccccagagagcctaaagcaccagagaagaataatgctctcactcaagctgac 2475  
DB 845 TTGACGGGCCCCAGAGCTTAAGCAGCAGCAAGAAATCAATGCTCTCACTACACCGTAC 904  
QY 2476 ccaacttctcagtgtagcagc 2535  
DB 905 CCAACTTTCTAGTGTCCACGCGCCACACACCTCCGACATCCACACCATCACCACCT 964  
QY 2536 gcttcctctcccaagcagtgatctgtatcttgatctatcttcttcttgatgataatg 2595  
DB 965 GCTTCTCTTCCACACGATGATCTGATCTTATGTTATTTATTTTGTGATGATGATG 1024  
QY 2596 aacataataaaatttcatcttgagaattctcactatgtaactgtaataagacagct 2655  
DB 1025 AACATATATATAAATTTTCATTGTGAGATTTCTCAATTTGATATTAATATGACAGATT 1084  
QY 2656 tggaaactgtctgagacgacttcaataataatcaacgcgaagaagatcatatcagct 2715  
DB 1085 TGGAAACTTGTCTGAGACTGACTTATCAATTAATCAACGACAAAGATATATCATATGT 1144  
QY 2716 gtagtgtagtagaatttattcatctgactgaacccaagagagcttcagagatgacaat 2775  
DB 1145 GTATGTGTTAGACATTTTATTTATTTGATGACTAACCCAGGACAGTTTCAAGTGAAT 1204  
QY 2776 tgtgtgcccctgtgtcagctgaaacagctcctggaacttcaaaacttgataagctc 2835  
DB 1205 TGTGTGCCCTCTGCTTACGCTGAAGACAGTCTCTGACCTTTCAAAAACCTTGAATAGCTC 1264  
QY 2836 ccacagttgataaaattggaacatttaggaattttaaacttagatgataatctgtctc 2895  
DB 1265 CCACAGTTGATATAATTTGACCAATTTAGGAATTTTAACTTATGATATCTTTGGTTCC 1324  
QY 2896 attttatcatatttatt 2955  
DB 1325 ATTTTATTTTCAATTTTATTTTGTATGCAAAACAGAGACTTAATGAACTTTATCTCT 1384  
QY 2956 gttttaagaattataaaacattgtgtaactatataatgagctctgagagactagct 3015  
DB 1385 GTTTTAAGATTAATTAATAAATTTGATGATCTATATATATGAGCTCTGAGAGATTAGCT 1444  
QY 3016 ttcaactaactaagagatgataatcagctgagctgagctgagctgagctgagctgagctgagct 3075  
DB 1445 TTCACTACACTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1504  
QY 3076 caagagtgctgatactgtttaatacatctccatgaagcgtgaagaagaatgaacagctt 3135  
DB 1505 CAGAGTCTGCATATCTGTTAATTAATCTCATTTAGGCTGAAGAAGATGACCTACGCTT 1564  
QY 3136 ctgatacagctgtgtgtcttcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3195

```
|||||
Db 1565 CTTGATACAGCGTGTGCTGTTTGGATGTTGCTACTGTACACAAAGTGTGCACGTA 1624
Qy 3196 ggcctcgtcgtgtgtccgtatggaacacctgtgtaaccccgagtgtaagtaactgtctca 3255
Db 1625 GCGCTGCGTGTGTGTCGATGGAAGCTGTAGCCCTGCCAGTTAACTACTGCTTCCA 1684
Qy 3256 ttcatgttcaatcgtgaatttctcctcccaatgaatgaatgaactaaactgaattgt 3315
Db 1685 TTCAATTGTACGCTGGAATTTTTCGCCCATGGAATGAAGTAAACTAAGTGTGTGT 1744
Qy 3316 catcaataatgttaataactaaataaaaaa 3348
Db 1745 CATCAATAAATGTAATACTAAAAA 1777

RESULT 2
BE867390 1018 bp mRNA EST 27-SEP-2000
LOCUS 601442470P1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3846689 5',
DEFINITION mRNA sequence.
ACCESSION BE867390
VERSION BE867390.1 GI:10316166
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1018)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
CNA Library Preparation: Life Technologies, Inc.
CNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L10M546 row: 1 column: 18
High quality sequence stop: 654.
Location/Qualifiers
1. 1018
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3846689"
/clone_11b="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-Sport6; Site_1: Nofl;
Site_2: Salt; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT 295 a 267 c 262 g 194 t
ORIGIN

Query Match 20.9%; Score 701; DB 136; Length 1018;
Best Local Similarity 90.0%; Pred. No. 6.8e-147;
Matches 809; Conservative 0; Mismatches 80; Indels 10; Gaps 5;
```

```
|||||
Db 121 AAGGAACGGGCACCAAGAAGGTGGCCAAAGCAATGCACCGACGAAACATGTGGAGAA 180
Qy 1302 tccctgtgttcaaatcccgagcgcaagcccaaacccgactcaagtcagagaga 1361
Db 181 TCCTTGTTTCAAAGTCCCGCAGCGGCGACCCCAACCCGACTCAAGTACAGAGACA 240
Qy 1362 agacaccccaataagaacacagagagatggaagaagaatcaacttttgaacctgtc 1421
Db 241 AGACACCCCAATAAAGAACAGGAGTGAAGAAAGTAACTTTTGTGAACCTGGCTGTG 300
Qy 1422 gggatggaataatgagactagtaataagagagatgaatgaatgcttactcaagtatc 1481
Db 301 GGGATGAATAATGGAGCTAATAATAAGAGATGAGTTCAGATGCTTATTAAGTCATC 360
Qy 1482 agcagctgcctcgtgaattcttcccaatgctgcccgaagtcgcccagagtcgaagata 1541
Db 361 AGCAGCTGCTGCTGGAATTTCTCCATGTGTCCCGAGGTCCCGACGCTGAGAGATT 420
Qy 1542 gtcaagacatcacacccaagaatttaccagagcagctcgaatctctgccaagcgca 1601
Db 421 GTCAAGGACATCACACCAAGATTTTACCAGGCGAGCTCGAATCTGCGCAAGGCCACGG 480
Qy 1602 taactgccaatgtagccgagagtggtgtatggtggggaacctgcgccacagcgagaca 1661
Db 481 TAACTGCCATGATGATGACCGGAGAGTTGCTGATGGGGGCACTCGCCACAGCCGAGACA 540
Qy 1662 tttaagaataatcctctcctcagagcagctacccatgagactctcaagagaacctgt 1721
Db 541 TTTTAAAGATATACATCTTCTTAAAGGCGACGTACCCATGAGCTCTCAGAACCTCTTG 600
Qy 1722 agcaactggaactatcttccagagtcagagagatccaggttgaatacaagaacttccca 1781
Db 601 AGCAACTGACATATCTTCCACAGTCCAGGAGTTCAGGTTGAATACAAAGACTTCCCA 660
Qy 1782 aaacaacaagaacgaattgtatct---cttaaatgtctctcctcagccactctga 1838
Db 661 AAAATCACCACCAAGCAAGATTCGATATCTCTTATCATATGCTCTCTGACCCACTGTGAT 720
Qy 1839 taagcatgtagtgcgaagagatgtgagtgctctgca---tgatagtgctgctggaaca 1895
Db 721 CCACATGATGATATGCGACAGGTGTGGGAATCTGTGCATATGATCATGGTGCTGGAC 780
Qy 1896 tcttaag-ttgcgtctcagatgtgacacaca--gtacagagatgccaagaacagga 1953
Db 781 TCTTTAAGATGTGCTGGCGAAGTGAAGCCACCAAGGTCAAAAATTTCCAGAAAGGGA 840
Qy 1954 cggaccaatgtctgtgtgtgg--agtgctgaaccttctgtgccaatgaaccattata 2010
Db 841 GGACCCAGTGTGCTGTTGGCAGACAGTCTCGAACCAATTTGGGCTGAACCCATTAA 899

RESULT 3
BE889880 1036 bp mRNA EST 29-SEP-2000
LOCUS 601512132P1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3913590 5',
DEFINITION mRNA sequence.
ACCESSION BE889880
VERSION BE889880.1 GI:10347645
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1036)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
CNA Library Preparation: Life Technologies, Inc.
```



[illegible]

QY	1089	aatatgagccaaagggatccaatccgatccgatctagccagatgagcccaatccagaaagccaaaagaag	1148
Db	1	AATATATGCCAGGGGGATCATATCCGATTATACCCAGCTGGCCCAATCCAGCAGGCAAAAAAGG	60
QY	1149	aaaagaagccaaagatcacacgcctccctcacagaagcgaagcctcccgcccgccgaaggaagtttg	1208
Db	61	ACAAGAGCGCAGACAGTACACGCTCCTCACAGAGCAGAGGCGTCCCGCGCGCAGGGAGTTTG	120
QY	1209	tgaatgcaggtgtaaggtatgynaacccaactgtcagaagaagaaagggcaaccaagaagatgtg	1268
Db	121	TGATCAGCAGGTGAAGGTTTGAAACCACTGTGAGAGAAACGGGCAACCAAGAAAGGTGG	180
QY	1269	ccaaagcgaatbtcaacccgaagaacatgctgcggagatccttggtttttaaagttcccgcaagcgcc	1328
Db	181	CCAAACGGCATACACCCAGAAACATGCTGGAGATCTTGTTTCAAAATGCCGAGGCGC	240
QY	1329	agcccaaccaaacccgcactcaagtacatagaggaagaacacataaagaacccaagggatg	1388
Db	241	ACCCACCAACACCCGCACTCAAGTGTAGAGGAAGAACACCCATAAAGAAACCA -GGGATG	299
QY	1389	gaagaagaatlaaccttlltgaacctggtccttgggatgtgaataatggaactagtataaag	1448
Db	300	GAGAGAAAGTAACTTTTGTGAACTGTGCTGTGGGGATGAAAAATGGGACTGTATATAAG	359
QY	1449	agatgaagtacagagatgctcttaactaaatcaagcagctgcctgcgtgaattcttccca	1508
Db	360	AGGATGAGTTTCAAGATGGCTTATCTTAAGTCATCAGCAGCTGCTGTGGATTTCTTCCCA	419
QY	1509	tggatgcccgagatctgcgccagagctgtgaagatgagtcaagaacatcacaccaagaattta	1568
Db	420	TGGTGCCCGAGGTGCGCCAGGCTGTAGAGATGTATCAAGACATCACACCAAAATTTTA	479
QY	1569	ccaaggcagctccggaatctctgcgaagggccaaogtbaactgcgatgatgcccgaagatgtg	1628
Db	480	CCAGGACACACTCCGAATCCTGTCCAGGCGCAGGTAACTGTGCATGATTAACCCGAGAGTTGT	539
QY	1629	tgtatggagggacccctcgcccaacagccgagacattttaa -gaataacatctcttcagcg	1687
Db	540	AGTATGGGGGCACTCTGCGCCACAGCCGAGATCATTTTAAAGAATMAATCATCTCTTCAAGC	599
QY	1688	caagtaaccatlgacactctcaagaaacccctctgagcaactgtgactatcttccagaagtc	1747
Db	600	CACGTAACCCCATAGGA -CTCTCACGAGATCCCTGTAGCAA -TGGACTATCTTCCAGAGTCC	657
QY	1748	caggatattccaggttgaataacaaagaacttcccaaaaaacaaagaagaatgtgatct	1807
Db	658	CAGGATATCCGGGTATACAAAGACTTTTCCAAAAAACAAGAAATTTGGTTCTATCA	717
QY	1808	ctatac 1813	
Db	718	ATGCTC 723	
RESULT	6		
LOCUS	BE300155/c	829 bp	mRNA
DEFINITION	600944602n1 NIH_MGC_17	Homo sapiens	cdna clone IMAGE:2960593 3',
ACCESSION	BE300155		
VERSION	BE300155.1	GI:9183903	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 829)		
JOURNAL	NIH-MGC <a href="http://www.ncbi.nlm.nih.gov/MGC/">http://www.ncbi.nlm.nih.gov/MGC/</a> .		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Tel: (301) 496-1550		
	Email: Robert_Strausberg@nih.gov		



Tissue Procurement: ATCC  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: L1CM54 row: d column: 02  
 High quality sequence start: 41  
 High quality sequence stop: 759.  
 Location/Qualifiers

## FEATURES

source

1. 829  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2960593"  
 /clone\_1ib="NIH\_MGC\_17"  
 /tissue\_type="rhabdomyosarcoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: muscle; Vector: pOTB7; Site\_1: EcoRI;  
 Site\_2: XhoI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
 for average insert size 1.8kb. Library constructed by  
 Ling Hong in the laboratory of Gerald M. Rubin (University  
 of California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."  
 BASE COUNT 190 a 200 c 226 g 213 t  
 ORIGIN

Query Match 18.2%; Score 610.4; DB 106; Length 829;  
 Best Local Similarity 99.7%; Pred. No. 1.4e-126;  
 Matches 622; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

378 cagagttttaccacttaaccttcagaactgaacaagaagaatgttctcctgaac 437  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 649 CAGAGTTTATTAACCTTAACCTTCAGAACTGAAACAAACAAATGTTCTCG-AC 591  
 438 gccctctttttaaagaagaacataccctactctagaactaatgcactgtgcatga 497  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 590 GCCCTCTTTTAAAAAAGAAAGCATTAACCCCTTAAGTAACTAATGCACTGTGATGA 531  
 498 aacttggaaaaaacaactgataagcctgttgacccttactctcgatgcagtlccact 557  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 530 AACTTGGAAAAAACCATATGTAACCTGTGACCTTACTCTCGATGCATCCACCT 471  
 558 ataactacaacttgagagagtgcttattccccgagtgacttttaaccaattccagttc 617  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 470 ATTAACCTACAACTGAGAGAGGTGCTTATCCCGAGGTACTTTTACCAATTCACAGTTTC 411  
 618 caacttacttatacaagtggaaacttctgtgtgagagagacaacttaatgcaagaagaa 677  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 410 CACCTTACTTATCAAGTGAACCTTCTGTGTGGAGACAGCAATTTAATGCGAAAGGAA 351  
 678 agacaagaagagctgcgaacaacgactgctgcgcaagaagcgttgaagatccctgcagaatg 737  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 350 AGACAAGAGAGGCTGGGAACACAGATGCTGCTGCCAAACGTTGAGAGATCCTGCACAATG 291  
 738 agccccctgcagaagagctgtgaggtgtaatggaagaatccgaaagaagaatctcaata 797  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 290 ACCCCCTGCCAAGAGAGGCTGAGAGTGAATGGAACATCCGAAGAAATAATCTCAATG 231  
 798 aatctgaataatgaatgaaggtttgagatgtgcaacttaaaoggaactgtgctgtgaatttg 857  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 230 AATCTGAATAATGAAGTCAAGGTTTGAATGACCTTAACGGAACCTGCTGTAATTTTCG 171  
 858 aggtggccccggagagatggtgcccaaccacatgaagaacttgytgcacaaggtttcggttg 917  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 170 AGGTGGCCCGGAGATGTGCCCAACCCACATGAGAACTTTGTGACCAAGGTTTCGTTG 111  
 918 gggaggttctgtggggagagtgaaagggaaaaagcaagaattctcaaaagaatgtccgca 977  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 110 CGGAGTTTGTGGGCGGAGAGTGAAGGAGAAAGCAAGAAATTTCAAGAAATAATGCGCCCA 51

QY 978 tagctgttcttgagagcgtgaaga 1001  
 ||||||||||||||||||||||||  
 Db 50 TACCTGTCTTCGAGAGCGTGAAGA 27

RESULT 7  
 BE379687 821 bp mRNA EST 21-JUL-2000  
 LOCUS BE379687/c  
 DEFINITION 601159257T1 NIH\_MGC\_53 Homo sapiens cDNA clone IMAGE:3510806 3',  
 mRNA sequence.

ACCESSION BE379687.1 GI:9325052

VERSION BE379687

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 1 (bases 1 to 821)  
 NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: ATCC

cDNA Library Preparation: CLONTECH Laboratories, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: L1CM196 row: a column: 15  
 High quality sequence start: 19  
 High quality sequence stop: 755.  
 Location/Qualifiers

## FEATURES

source

1. 821  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3510806"  
 /clone\_1ib="NIH\_MGC\_53"  
 /tissue\_type="carcinoma, cell line"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: bladder; Vector: pDNR-LIB (Clontech);  
 Site\_1: SfiI (ggcgctcgccgc); Site\_2: SfiI (ggcctatggcc  
 ); Double-stranded cDNA was prepared from cell line RNA.  
 5' and 3' adaptors were used in cloning as follows: 5'  
 adaptor sequence: 5'-CACGGCCATTAATGAC-3' and 3' adaptor  
 sequence: 5'-ATTCTGAGAGCGGAGCGGCGACATG-dT(30)BN-3'  
 (where B = A, C, G, or T). Average  
 insert size 1.55 kb (range 0.9-4.0 kb). 15/15 colonies  
 contained inserts by PCR. This library was enriched for  
 full-length clones and was constructed by Clontech  
 Laboratories (Palo Alto, CA)."  
 BASE COUNT 291 a 154 c 150 g 226 t  
 ORIGIN

Query Match 18.1%; Score 606.4; DB 107; Length 821;  
 Best Local Similarity 96.8%; Pred. No. 1.1e-125;  
 Matches 672; Conservative 0; Mismatches 16; Indels 6; Gaps 5;

2651 cagttggaaactgtctgagactgtatc-aataatcaaccgacaagaatataat 2709  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 820 CACAGTTGGAACTGCTGAGAGGACTGTATGAAAAATCTAACGACAAAGATCATAT 761  
 2710 ccatgtgtatgtgttagaactttat-ttcaattgaaccacagagatctcagta 2768  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 760 CATGTGTATGTGTAGACATTTTATGTTCATTGACTTAACCCAGAGAGTTTCAGTGA 701  
 2769 tgcgaattgtgtgcccctgtgttcagctga-aacagtcctgtgac--ttcaaaaccttg 2825  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 700 TGCAAATTTGTGCTCCCTCTGCTTACAGTGAAGACAGTCTGAGAGCTTTCAAAAACCTTTG 641  
 2826 aataagttcccccagttgtataaatgtgacaaatttagagaaatttaacttagatgagtc 2885

```

Db 640 AATAAGTCCACAGAGTGTATTAATGACAAATTAGAAATTTAACTTAGATGATC 581
Oy 2886 atttggtccaat-ttatattcaatttatttcttgtaataacagagacttaataa 2944
Db 560 AATTGGTTCATGTTTATTTTCATTTTATTTTGTATGCAACAGAGACTTAATGA 521
Oy 2945 cttaacatctgtcttaagaattataaacaacatctgtatcatatcatatgctctg 3004
Db 520 CTTTATCTCTGTCTTTAAAGATTTATTAACATTTGTGTATCTATCATATGCGCTTGG 461
Oy 3005 aggaactagcttccactacactacagatatgatcatcatgtaagctacataaacttga 3064
Db 460 AGGACTTAGCTTTCACACTACACTACAGATATGATCTCCATGATGCAATTAACCTGGA 401
Oy 3065 gagtgatttccagagatgctgtatctgttaatactatcatatgagctgtaagaagt 3124
Db 400 GAGTGAATTTCCAGAGAGCTCGATCTGTAAATACATCTCCATTAAGGCGTGAAGAAAT 341
Oy 3125 gacctagcttctgtatatacagctgtgttgccttgatgttggttacctatcacagaagt 3184
Db 340 GACCTACGTTCTGTATACAGCTGTGTGCTTTTATGTGTGTACTGTACACAGAACT 281
Oy 3185 gttgcaactgaagctctgctgtgtgtccgtatgaaacactgtgtagccctgcgaagtaag 3244
Db 280 GTGTGACAGAGAGCTGTGCGTGTGCGTATGAAAGCGCTGTAGCCCTGCGAGTTAAG 221
Oy 3245 tactgtccatcatctgttctacgtgtgaatttctcccatgtaagtaagtaaac 3304
Db 220 TACTGCTTCCATTCATGTTTACGCTGGAATTTTCTCCCATGGAATGTAGTAAACT 161
Oy 3305 taagtctgtcatcaataaagtgaactaa 3338
Db 160 TAAGTGTTCATCATATAATGTAATACATAA 127

```

```

RESULT 8
BE380123 609 bp mRNA EST 21-JUL-2000
LOCUS 601159257F2 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:3510806 5',
DEFINITION mRNA sequence.
ACCESSION BE380123
VERSION BE380123.1 GI:9325488
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 609)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at: image.lnl.gov
Plate: L1CM196 row: a column: 15
High quality sequence stop: 606.

```

## FEATURES

```

Source
1..609
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3510806"
/clone_lib="NIH_MGC_53"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (TI phage-resistant)"
/notes="Organ: bladder; Vector: pDRK-LIB (Clontech);
Site_1: Sfil (ggccgcctggcc); Site_2: Sfil (ggccattatggcc

```

```

); Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCATATGAGGC-3' and 3' adaptor
sequence: 5'-ATTCTAAGGCGCCAGGCGGCGGCACATC-dt(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.55 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
BASE COUNT 178 a 125 c 96 g 210 t
ORIGIN

```

```

Query Match 18.0%; Score 604; DB 107; Length 609;
Best Local Similarity 100.0%; Pred. No. 3,8e-125;
Matches 604; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 2395 agccatgccagagagagagccttgaccgccccagagactaaagaccagagaaatca 2454
Db 1 AGCCATGCCAGAGAGAGAGCGCTTTGACCGGCCAGAGCTAAAGCACAGAAATCA 60
Oy 2455 atgtctctactcaagcgtagcccaactttctagtgtgtccaaagcccaaccctctgc 2514
Db 61 ATGCTTCTACTACAGCGTAGCCCACTTTCTAGTGTGCCAGGCCCCACACCTCTGCG 120
Oy 2515 agtaccacacacacacacacacagcttctctccacacagtgatctatcttattcat 2574
Db 121 AGTACCACACACATACCACTGCTTCTCTTCCACAGTGAATGATATTTCTAGTTTCA 180
Oy 2575 taattctcttgatgatataagacatataaatttcaatttgaagattctcaattgt 2634
Db 181 TATTTCTTTTGATGATATGATGACATATATAAATTTTCATTTGAGAAATTTCTCAATTGT 240
Oy 2635 atctagttaaaagacagcttggaaactgtctgaagctgaacttacaataactaac 2694
Db 241 ATCTAGTTAAATAGCAGCAGTTTGGAACTTGTCTGAGACTGACTTATCAATAATCTAAC 300
Oy 2695 cgacaaagatcatcatcatctgtatgtgttgtagacatttcaatgtaactaacag 2754
Db 301 CGACAAAGATCATATCATATGATGTATGTGTATGACATTTTATTTTCATGACTAACCCAG 360
Oy 2755 gacagttcagtgatgacaaatgtgtccctctgtgtcgaagctgaacagctctgactt 2814
Db 361 GACAGTTTCAGTGAATGCAAAATGTGTGCGCTGTGCTGACGTGAACAGTCTCGACCTTT 420
Oy 2815 caaaaacttgataaagctctccacagttgtatataattggacaatttggaaattttaa 2874
Db 421 CAAAACCTTGAATAGTCTCCACAGTTGTATTAATTGGACAATTTAGGAATTTTAAAC 480
Oy 2875 tttaagtatcatctgtgtccatttctatcttcaatttcttgaatgaaacagga 2934
Db 481 TTTAATGATCATATTTTGTGCTCCATTTTATTTATTTTATTTTATGCAACAGAGA 540
Oy 2935 cttaaatgaacttgatctctgttcttaagaattataaacaactgtgtatcatat 2994
Db 541 CTTAATGAACCTTGTATCTGTGTTTAAAGATTATTAACAACTGATATCTATATCAT 600
Oy 2995 atgg 2998
Db 601 ATGG 604

```

```

RESULT 9
BE613972 686 bp mRNA EST 24-AUG-2000
LOCUS 601504093F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3905783 5',
DEFINITION mRNA sequence.
ACCESSION BE613972
VERSION BE613972.1 GI:9895569
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE 1 (bases 1 to 686)  
 AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert\_Strausberg@nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LHC700 row: j column: 24  
 High quality sequence stop: 675.

FEATURES  
 Location/Qualifiers  
 1..686  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3905783"  
 /clone\_lib="NIH\_MGC\_71"  
 /rissue\_type="leiomyosarcoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.1 kb."

BASE COUNT 226 a 156 c 196 g 108 t  
 ORIGIN

Query Match 18.0%; Score 603.8; DB 110; Length 686;  
 Best Local Similarity 97.1%; Pred No. 4.3e-125;  
 Matches 669; Conservative 0; Mismatches 12; Indels 8; Gaps 5;

696 aacacagatctgctgcaaaagcgtltagagatctctgagaatgagccctgccaagagc 755  
 1 AACACATGCTGCTGCGCAAAAGCTTGAGATCCTGCAGATGAGCCCTGCCAGAGAGC 60  
 756 tggaggtgagtggaagagatccgaaagaatactcaataaataagtaagtaag 815  
 61 TGGAGGTGATGAGAGAGATCGAAGAAATCTCAATAATCTGAATAATGAGTCAAG 120  
 816 tttttagagttgacttaagaagacttgcgttgaatttcgagtgagtgagccggagagtg 875  
 121 TGTTTGAGATGACTTAAACGGAAGCTGCTGTGAATTTGAGGTGCGCCGAGAGAGTG 180  
 876 gccacccacatgaagaacttctgtaaccaagtttcggttgggagttgttgaggag 935  
 181 GCCACCCACATGAGAGAGC-TGTGTACCAAGTTCGCT--GGGAGTGTGCGGAGAG 237  
 936 gtgaagggaaagcaagaagattcaagaagaaatgcgcacatagctgttcttgagagc 995  
 238 GTGAGAGGGAAGCAAGAGATTTCAGAGAAATGCCGCATAGCTGTCTTGAGAGAGC 297  
 996 tgaagaagttaccgcccctgctgagttgaagagtaagccttagaatcaagaagaaa 1055  
 298 TGAAGAAGTTACCGCCCTGCTGCTGAGTTGAAGAGGTAAGCCTTAGAATCAAAAAGAAA 357  
 1056 caaaacccatagtcacagcaccagacagccagaatatagtccagggatcaatccgalt 1115  
 358 CAAAACCCATAGTCAGCCACAGCAAGCCCAATATGSCCA--GGGATCAATCCGATT 416  
 1116 gccgaatgcccagatccagcaggaagaaaggaagagagccagagttacagctctca 1175  
 417 GCCGACTGCCCCAGATCCAGCAGCAAAAAAGAGAGCCAGATACACCTCTCTCA 476  
 1176 caagaagaagccctccgcccagcaggaagtttgaatgagtgagtgagtgagtgagtgag 1235  
 477 CAGAGCGAGAGCCCTCCCGCCGCGCAGGAGAGTTTGTGATGCGAGGTGAGAGTTGAGAAC 536  
 1236 ctgcagaagaagcggcaccacaagaaggtggcgaagcgcaatgcagccgaacaatgc 1295

|||||  
 Db 537 CTGCAGAGAGAGAGGAGCCACACAGAGGTGGCCAGGCGCAATGACGCGAAGACTTGC 596  
 1296 tggagatcccttgg-tttcaagttcccgagcggcagccacacacccgacccaagtca 1354  
 597 TGGAGATCCTGTTGTTTCAAGTCCCGCAGGCGCAGCCACCAAA---CCGACTCAAGTCA 653  
 1355 gaggagaagaacaccataaagaacacag 1383  
 654 GAGAGAGAGACACCCATTAAGAAAGAGAGG 682

RESULT 10  
 AV707406 674 bp mRNA EST 09-OCT-2000  
 LOCUS AV707406 ADB Homo sapiens cDNA clone ADBCG03 5', mRNA sequence.  
 DEFINITION AV707406  
 ACCESSION AV707406.1 GI:10724671  
 VERSION AV707406.1  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 674)  
 AUTHORS Peng,Y., Song,H., Huang,Q., Huang,C., Gu,Y., Yang,Y., Gao,G., Xiao,H., Xu,X., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R., Chen,J., Chen,Z. and Han,Z.  
 Homo sapiens cDNA ADB clones  
 Unpublished (2000)  
 TITLE Chinese National Human Genome Center at Shanghai  
 JOURNAL Contact: Zeguang Han  
 Chinese National Human Genome Center at Shanghai  
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
 201203, P. R. China  
 Tel: 86-21-50801919(ex.45)  
 Fax: 86-21-50801922  
 Email: hanzg@chgc.sh.cn

FEATURES  
 Location/Qualifiers  
 1..674  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="ADBCMG03"  
 /clone\_lib="ADB"  
 /rissue\_type="Adrenal gland"  
 /dev\_stage="Adult"  
 /lab\_host="SOLR"  
 /note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 197 a 148 c 113 g 215 t 1 others  
 ORIGIN

Query Match 18.0%; Score 602.2; DB 38; Length 674;  
 Best Local Similarity 98.4%; Pred. No. 9.7e-125;  
 Matches 618; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

2352 tctcaagctgctcactcacaacacatttggagcaaaccttggagcgcacatgcagaagag 2411  
 1 TCTCAAGCTGCTCCTCCTCAGACACATTTGGAGCAAAACCTTGAGACGCGCATGCGACAGAGAG 60  
 2412 gcccttgacggccgcagagcctaaagcacagagaaatcaaatgcttctactacagc 2471  
 61 GCCTTGACCGGCGCCAGAGCTAAAGCACCAGAGAAATCAATAGCTTCTACTACAGCG 120  
 2472 tgaaccaactttcttagtgtgccaagggcccaaccactctcgaagtaaccaaccatcac 2531  
 121 TGACCCCAACTTTTCTAGTGGCCACGCGCCACACCTCTGCGAGTACCCACACCATAC 180  
 2532 cactgcttctcttccaagagtgatctgtattcttagttcatatttcttttgatga 2591  
 181 CACTGCTTCTCTTCCCAACAGTGTGATCTTCTTACTTTCAATTATTTCTTTGATTGA 240



Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL, send email to:  
 info@image.llnl.gov  
 Seq primer: -400P from Gibco

High quality sequence stop: 469.

## FEATURES

source

Location/Qualifiers

1..616

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:3278862"

/clone.lib="NCI.CGAP.Lu24"

/tissue\_type="carcinoid"

/lab\_host="DH10B"

/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a

modified polylinker; Plasmid DNA from the normalized

library NCI.CGAP.Lu5 was prepared, and ss circles were

made in vitro. Following HAP purification, this DNA was

used as tracer in a subtractive hybridization reaction.

The driver was PCR-amplified cDNAs from a pool of 5,000

clones made from the same library (clones

1414920-1417991 and 1520904-1522439). Subtraction by Bento

Soares and M. Fatima Bonaldo.

Scores and M. Fatima Bonaldo.

BASE COUNT 179 a 153 c 139 g 141 t 4 others

ORIGIN

## Query Match

Best Local Similarity 17.8%; Score 594.4; DB 111; Length 616;

Matches 598; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1361 aagacacccataagaacccaggagatggaagaagaactcttttgaaactgctct 1420

DB 8 AAGACACCCCTAAGAAACCGGGGATGGAAGAAACTTTTGAACCTGGCTCT 67

QY 1421 gggagataaattgggactagtaataagagatgagatgagcttcttaagtc 1480

DB 68 GGGGATGAATAATGGACTACTAATAAGAGATGAGTTCAGAGTCCCTTATCTAAGTCAT 127

QY 1481 cagcagctgctgctggaattctcccatggtgcccgaggcgccagcgtgtagagt 1540

DB 128 CACACAGCTGCTCTGGAATTTCTCCCATGTGTGCCAGGTGCGCCAGGCTGTAAGATT 187

QY 1541 agtcaagagatcacaccaaagatttaccaggagcgtccgaactcgtccaaagcag 1600

DB 188 AGTCAAGGACATCACACCAAAAGATTTTACACAGGCGACCTCGAATCTGCCAAGGCCACG 247

QY 1601 gtacatgcatagtagccgaaggtgtgtataggggcgacccctcgccacagccgagacc 1660

DB 248 GTACTGCTCCATGATGCCAGAGGTTGTATGGGGGACCTCGCCACAGCCGAGAGCC 307

QY 1661 atttlaaagataacatctcttcagcgcaagtaaccatgagactcttcacagagacctct 1720

DB 308 ATTTTAAAGATTAACATCTCTTCAGGCCAGTAAACCCATGAGACTCTTCACAGAGACCTCT 367

QY 1721 gagcaactgactatcttccagagtcagaggtatccagggttgataataaagactcccc 1780

DB 368 GAGCAACTGACATCTTTTCCAGAGTCCAGGATTTCCAGGTTGAAATCAAAAGACTTCCCC 427

QY 1781 aaaaacaagaagaagatgtatctcttatactatgctccctcttgaagcaacctctatc 1840

DB 428 AAAAACAACAAGAAGCAATTTGTTATCTTATCAATTTGCTCTTCANCCACCTCTGATC 487

QY 1841 agccatgtagtcgcaagagatgtgagagtcctgcatgatalatgctgctgcaaatctta 1900

DB 488 AGCCATGTTATCAGCAAGATGTGAGTCCGATGATATGCTGGCGTAACATCTTTA 547

QY 1901 aagttgctgctgagtcgagcaaaagatcacagagatgcacaagaagaaagcagca 1960

DB 548 GAGTTGCTGTGTAATGTGACCAAAANTACAGAGATGCCAAGACANGANACAGACCA 607

QY 1961 atgtctgt 1968  
 |||||

DB 608 ATGCTCT 615

## RESULT 13

AM162002/c

LOCUS

DEFINITION

AM162002

594 bp mRNA

EST

09-NOV-1999

au72c03.x1

Schneider fetal brain 00004 Homo sapiens cDNA clone

IMAGE:2781796 3' similar to TR:095793 095793 STAU6N PROTEIN. ;

RNA sequence.

AM162002

GI:6301035

EST.

KEYWORDS

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 594)

Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,

Krivan, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin

, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theisling, B.,

White, Y., Wylie, T., Waterston, R. and Wilson, R.

WashU-NCI human EST Project

Unpublished (1997)

Other ESTs: au72c03.y1

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -400P from Gibco

High quality sequence stop: 464.

Location/Qualifiers

1..594

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2781796"

/clone.lib="Schneider fetal brain 00004"

/sex="male"

/tissue\_type="frontal lobe"

/dev\_stage="5 months post-conception"

/lab\_host="DH10B"

/note="Organ: brain; Vector: pBluescript SK (Stratagene);

site:1: SstI; site:2: XhoI; Double-stranded cDNA was

prepared from human fetal brain tissue. 5' and 3'

adaptors were used in cloning as follows: 5' adaptor

sequence: 5'-GAGAGAGAGAGAGCTCAGAGATCTTAATAATTAATCCCCCCCCC-3'

and 3' adaptor sequence:

5'-GAGAGAGAGAGCTCAGAGATCTTAATAATTAATCCCCCCCCC-3'

insert size estimated for >0.5 kb inserts and has an average

insert size estimated at 1.2 kb. This library was

constructed using the CAP-trapper method for full-length

enrichment and has not undergone amplification. Library

was constructed by Dr. Claudio Schneider (UNCIB-Area

Science Park, Trieste, Italy).

BASE COUNT 130 a 152 c 123 g 188 t 1 others

ORIGIN

## Query Match

Best Local Similarity 17.6%; Score 588.2; DB 40; Length 594;

Matches 590; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 455 gaagcataaccctactgtgaactaataatgcatgcatgaacttggaataaaca 514

DB 594 GAAAGCATAAACCCCTACTGTGAAGATTAATGCTGCTGATGAAGAACTTGGAACCA 535

```

"organism"="Homo sapiens"
"db_xref"="taxon:9606"
"clone"="GKCEFE08"
"clone_1ib"="G8C"
"tissue_type"="hepatocellular carcinoma"
"dev_stage"="Adult"
"lab_host"="SOLR"
"note"="vector: pBluescript sk(-); Site-1: EcoRI; Site-2

```

Email: Robert\_Strausberg@nih.gov

Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNLN at:  
<http://image.llnl.gov>  
 Plate: LNCM545 row: k column: 02  
 High quality sequence stop: 705.

## FEATURES

Location/Qualifiers  
 1..749  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3846265"  
 /clone\_lib="NIH\_MGC\_65"  
 /tissue\_type="adenocarcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: colon; Vector: pCMV-SPORT6, Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
 Average insert size 1.8 kb. Library constructed by Life  
 Technologies."

BASE COUNT 244 a 166 c 223 g 116 t  
 ORIGIN

Query Match 16.9%; Score 565.4; DB 110; Length 749;

Best Local Similarity 95.1%; Pred. No. 1.8e-116;  
 Matches 714; Conservative 0; Mismatches 21; Indels 16; Gaps 12;

```

QY 677 aagacaagaacagctgcgaacaacgatctgtccaaa-gcgttgaagatcctcagaa 735
    |||
DB 1 AAGACAAAGACAGGCTGCAGAACACGATGCTGCGCAAAAGCGTGAAGATCCTGCAGAA 60

QY 736 taagccctgcgaagagctggaatgaatgaagaatccgaagaagaatctcaa 795
    |||
DB 61 TGAAGCCCTGCAGAGAGGCTGAGGTGAATGAGAGAAATCCGAAAGAAATCTCAA 120

QY 796 taatctgaataaagtcgaagtgtt-gagatcacttaacggaacttgctgtaatt 854
    |||
DB 121 TAAATCTGAATAATAGTCAGTGTGAGATTCGACTTAACGGAATTCCTGTGATTT 180

QY 855 tcgaagctgcccggagagctgcccacccacatgaagaacttltgaaccaagtttcg 914
    |||
DB 181 TCGAGGTGCGCCGGAGAGTGGCCACCCACATGAAGAAC-TTGTGACCAAGTTTCGG 239

QY 915 ttgggaagtttgggggaagtgaaaggaagaaagcaagaattccaagaataatgcg 974
    |||
DB 240 TTGGGGAG-TTGTGGGGGAGGTGAAGGGAAGCAAGAGATTTCAAAAGAAAATGCCG 298

QY 975 ccatacgtctcttgaggagctgaagaagtltaacgcccctgcctgcagtgaacgagtaa 1034
    |||
DB 299 CCATACCTGTCTTGAAGAGCTGAAGAGTTACCG-CCCTGCTGCAGTTGAAGAGACTAA 357

QY 1035 agcctagaatcaaaaagaacaaacccaatagtcgaagccacagacaagccagaatatg 1094
    |||
DB 358 AGCCTAGAATCAAAAAGAAAACCAATAGTCAAGCCACAGACAAGCCAGAAATATG 417

QY 1095 gccgaaggatcaatccgattagccgactgcccagatccagcagcgaataaaggagaagg 1154
    |||
DB 418 GCCA-GGGATCAATCCGATTAGCCGACTGCGCATTCAGAGCAAAAAGAGAGAGG 476

QY 1155 agccaagatacaacgctcccaagagcgaagcctcccgccgacgaggaatttgaatgc 1214
    |||
DB 477 AGCCAGATGACAGCTCTCTACAGAGCGAGG-CTCCCGCCGCCAGAGGAG-TTGTGATGC 534

QY 1215 aggtgaagtttgaaacccaactgcagaaggaagcagcaaac-aagaagctgcgaag 1273
    |||
DB 535 AGGTGAAGGAGGGAACACACTGCAAGAGGACGACCAACAAGAGGTGGCCAAG 594

QY 1274 cgcgaatgcagcagagaacatgctggaatccttggtt--caaagtcgcgacgcgcagc 1331
    |||
DB 595 CGCAATGCAAGCCGAGAACATGCTGGAGATCCTTGTTCAAAAGTCGCCAGGGGCGAGC 654

```

```

QY 1332 ccaccaaacccgcactcaagtlcagagagagaagacaccacataaagaacacgaggaatgaa 1391
    |||
DB 655 CAACAAACCGGACTC--AGTCAGAGAGAGAAGACACCCATTA--GAAACCGGGGATGGAA 709

QY 1392 gaaaagtaaccttttttgaaacctggtcttg 1422
    |||
DB 710 GACACGTAAACCTTTTGGAACTGGGCGGGG 740

```

Search completed: April 5, 2001, 00:30:32  
 Job time: 53128 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2001, 08:31:19 : Search time 701.87 Seconds  
(without alignments)  
1791.954 Million cell updates/sec

Title: US-09-316-048-7  
Perfect score: 3348  
Sequence: 1 acttcctgcgcggcgtcgcgg.....taatactcaaaaaaaaaa 3348

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 18783343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: N.Geneseq\_36:\*  
2: /SID6/gcgdata/geneseq/geneseqn/NA1980.DAT:\*  
3: /SID6/gcgdata/geneseq/geneseqn/NA1981.DAT:\*  
4: /SID6/gcgdata/geneseq/geneseqn/NA1982.DAT:\*  
5: /SID6/gcgdata/geneseq/geneseqn/NA1983.DAT:\*  
6: /SID6/gcgdata/geneseq/geneseqn/NA1984.DAT:\*  
7: /SID6/gcgdata/geneseq/geneseqn/NA1985.DAT:\*  
8: /SID6/gcgdata/geneseq/geneseqn/NA1986.DAT:\*  
9: /SID6/gcgdata/geneseq/geneseqn/NA1987.DAT:\*  
10: /SID6/gcgdata/geneseq/geneseqn/NA1988.DAT:\*  
11: /SID6/gcgdata/geneseq/geneseqn/NA1989.DAT:\*  
12: /SID6/gcgdata/geneseq/geneseqn/NA1990.DAT:\*  
13: /SID6/gcgdata/geneseq/geneseqn/NA1991.DAT:\*  
14: /SID6/gcgdata/geneseq/geneseqn/NA1992.DAT:\*  
15: /SID6/gcgdata/geneseq/geneseqn/NA1993.DAT:\*  
16: /SID6/gcgdata/geneseq/geneseqn/NA1994.DAT:\*  
17: /SID6/gcgdata/geneseq/geneseqn/NA1995.DAT:\*  
18: /SID6/gcgdata/geneseq/geneseqn/NA1996.DAT:\*  
19: /SID6/gcgdata/geneseq/geneseqn/NA1997.DAT:\*  
20: /SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:\*  
21: /SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:\*  
22: /SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3139.	93.8	3260	21	293327 Human stauferi cDNA
2	3076	91.9	3217	21	293323 Human stauferi cDNA
3	2987.2	89.2	3190	20	X80786 DNA encoding human
4	2961.6	88.5	3506	21	293322 Human stauferi cDNA
5	2926	87.4	3142	21	293324 Human stauferi cDNA
6	1255	37.5	2859	21	293328 Mouse stauferi cDNA
7	1107.2	33.1	1127	20	X39991 Prostate cancer as
8	798.8	23.9	946	20	X40002 Prostate cancer as
9	755.6	22.6	844	20	X40000 Prostate cancer as
10	753.2	22.5	773	20	X39992 Prostate cancer as
11	551	16.5	825	20	X40001 Prostate cancer as
12	501.4	15.0	794	20	X39993 Prostate cancer as

C	13	433.8	13.0	660	21	279981 Human colon cancer
C	14	412.2	12.3	727	21	280269 Human colon cancer
C	15	327.8	9.8	708	21	280703 Human colon cancer
C	16	286.6	8.6	305	16	T20075 Human gene signatu
C	17	119.8	3.6	300	20	217757 Human gene express
C	18	119.8	3.6	300	20	214226 Human gene express
C	19	115.2	3.4	769	20	217411 Human gene express
C	20	88.2	2.6	235033	19	V57926 Hereditary haemoch
C	21	88.2	2.6	237326	19	V57903 Hereditary haemoch
C	22	86.8	2.6	17327	14	O44278 Serglycin - proteo
C	23	85.8	2.6	6870	21	A34833 Human adenosine re
C	24	85.8	2.6	8055	21	A34834 Human adenosine re
C	25	85.4	2.6	228	19	V4914 Human Iga nephropa
C	26	85	2.5	13585	17	T11549 Tumour rejection a
C	27	85	2.5	56583	21	A35003 Human adenosine re
C	28	84.8	2.5	1354	19	V35619 Human transcriptio
C	29	84	2.5	84607	20	X90847 Human PACAP genom
C	30	83.8	2.5	9365	21	V50359 Human CD39-L4 geno
C	31	83.8	2.5	32367	19	V35620 Human SHOX (short
C	32	83.4	2.5	2529	21	A16662 Human secreted pro
C	33	83.4	2.5	138169	21	A34791 Human adenosine re
C	34	83.4	2.5	141589	21	A35005 Human adenosine re
C	35	83.4	2.5	141589	21	A35030 Human adenosine re
C	36	83.4	2.5	162450	21	Z86967 Retinoblastoma bin
C	37	83.4	2.5	17131	21	Z60888 DNA encoding a hum
C	38	83.2	2.5	17131	21	Z60888 Human PACAP genom
C	39	83	2.5	84607	20	X90847 SHOX gene exon VB
C	40	82.4	2.5	625	19	V35615 Human gene express
C	41	82.2	2.5	300	20	Z12542 Human secreted pro
C	42	82.2	2.5	1040	20	X79016 Human IL-1ra BAC c
C	43	82.2	2.5	2170	20	X03027 Human chromosome-1
C	44	82.2	2.5	3071	17	T11778 Human IL-1ra BAC c
C	45	82.2	2.5	11901	20	X02998 Human IL-1ra BAC c

## ALIGNMENTS

RESULT	1
ID	293327 standard; cDNA; 3260 BP.
XX	
AC	293327;
XX	
DT	04-JUL-2000 (first entry)
XX	
DE	Human stauferi cDNA.
XX	
KW	Staufen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW	HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW	acquired immune deficiency syndrome; protease; ; human; ss.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	Location/Qualifiers
FT	409..1311
FT	/*tag= a
FT	/product= Staufen protein
XX	
PN	CA2238656-A1.
XX	
PD	22-NOV-1999.
XX	
PF	22-MAY-1998; 98CA-2238656.
XX	
PR	22-MAY-1998; 98CA-2238656.
XX	
PA	(UYMO-) UNIV MONTREAL.
XX	
PI	Duchaine T, Wickham L, Cohen EA, Luo M, Desrosiers L;
XX	Moulard AJ;
DR	WPI; 2000-246924/22.

DR P-PSDB; Y83108.

XX Isolated mammalian or Caenorhabditis elegans staufen proteins useful  
PT for treating retroviral infections especially human immunodeficiency  
PS virus infections (HIV)

XX Claim 4; Figure 1b; 96pp; English.

CC Staufen is a RNA binding protein in which interacts with double stranded  
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA  
CC transport and localization. Mammalian SP contain multiple ds  
CC RNA-binding domains. Chimeric proteins comprising portions of  
CC staufen with proteins with RNase or protease activity can be used to  
CC treat virus infections. The RNase or protease activity of the fusion  
CC protein prevents proper maturation of the virus. Mammalian staufen  
CC proteins seem to recognise double stranded RNA structure rather than  
CC any sequence specific position. 2 bacterially-expressed fusion  
CC proteins used in an RNA-binding assay (his/HisP (human SP) and  
CC MBP/MSF (murine SP)) both proteins strongly bound double stranded  
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and  
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the  
CC treatment of RNA virus infections, especially those caused by  
CC retroviruses, in particular human immunodeficiency virus. This  
CC sequence is an alternatively spliced sequence to the one given in  
CC Z93322.

XX Sequence 3260 BP; 907 A; 795 C; 759 G; 799 T; 0 other:

Query Match 93.8%; Score 3139; DB 21; Length 3260;  
Best Local Similarity 97.3%; Pred. No. 0;  
Matches 3259; Conservative 0; Mismatches 0; Indels 90; Gaps 3;

QY 1 actctctgcggagctbctcgagcgctcttagcggtttgacgagcggtcggt 60  
DB 1 actctctgcggagctbctcgagcgctcttagcggtttgacgagcggtcggt 60  
QY 61 cctctcgcgctccggtctcttcttgacgctctcccccgcggcgcgcgccctc 120  
DB 61 cctctcgcgctccggtctcttcttgacgctctcccccgcggcgcgcgccctc 120  
QY 121 ctccacgagcactcgctcttctccctcctcgctccttcttcttcttctt 180  
DB 121 ctccacgagcactcgctcttctccctcctcgctccttcttcttcttctt 180  
QY 181 ctctctctccctctcgccgcaacgcccagacgcccggggagacgtccgaac 240  
DB 152 -----agc 154  
QY 241 agcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 300  
DB 155 agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 214  
QY 301 gtagctcgcgtcactcgcaactccacccccaagcagcagcagcagcagcagc 360  
DB 215 gtagctcgcgtcactcgcaactccacccccaagcagcagcagcagcagcagc 274  
QY 361 ccgataagcttgatatac-gagttatataaccacttaacccttcagaactgaacaaga 419  
DB 275 ccgataagcttgatatac-gagttatataaccacttaacccttcagaactgaacaaga 334  
QY 420 caacatgtctctggaagccctctttaaanaaagaacataaccctactgtagaac 479  
DB 335 caacatgtctctggaagccctctttaaanaaagaacataaccctactgtagaac 394  
QY 480 taatgtcactgtgcatgaaacttgaaaaaaaccaaagtataagcctgttgaccctact 539  
DB 395 taatgtcactgtgcatgaaacttgaaaaaaaccaaagtataagcctgttgaccctact 454  
QY 540 ctccgagtgagtcacactataacatacagagagagtgcttatcccccgaagtgact 599  
DB 455 ctccgagtgagtcacactataacatacagagagagtgcttatcccccgaagtgact 511

QY 600 ttaccacattccagctccacacttacttatacgaagtgaacttctgtgaggagcagc 659  
DB 512 ttaccacattccagctccacacttacttatacgaagtgaacttctgtgaggagcagc 571  
QY 660 aatttaatgcaaggaagaagaagaagaagaagaagaagaagaagaagaagaaga 719  
DB 572 aatttaatgcaaggaagaagaagaagaagaagaagaagaagaagaagaagaaga 631  
QY 720 tgaagatctctgcaagatgagccctccagagagcgtggaagtgaatggaagaatccg 779  
DB 632 tgaagatctctgcaagatgagccctccagagagcgtggaagtgaatggaagaatccg 691  
QY 780 aagaagaataatcctaataactctgaaataagtaagtgattgagattgcaactaaacga 839  
DB 692 aagaagaataatcctaataactctgaaataagtaagtgattgagattgcaactaaacga 751  
QY 840 acttgctgtgatttctcgaggtgagcccgagagatgagccacccacatgaagaacttg 899  
DB 752 acttgctgtgatttctcgaggtgagcccgagagatgagccacccacatgaagaacttg 811  
QY 900 tgaacaggttctcggttgggaggttctgaggagaggtgaaaggaaagaagaagaat 959  
DB 812 tgaacaggttctcggttgggaggttctgaggagaggtgaaaggaaagaagaagaat 871  
QY 960 caaagaataatgacgcacatagctgttcttgaggagcgtgaagaagtaccgcccctgctg 1019  
DB 872 caaagaataatgacgcacatagctgttcttgaggagcgtgaagaagtaccgcccctgctg 931  
QY 1020 cagttgaaacgagtaaaagcctagaatcaaaaagaataaaacccatagccaagcaga 1079  
DB 932 cagttgaaacgagtaaaagcctagaatcaaaaagaataaaacccatagccaagcaga 991  
QY 1080 caagcccaaatatgagcccgaggagatccgaatccgaattgagccagctgagccagcag 1139  
DB 992 caagcccaaatatgagcccgaggagatccgaatccgaattgagccagctgagccagcag 1051  
QY 1140 caaaaaaggaagaagaagcagaagtaacagctctccacaagaagcagcctcccgcgcgca 1199  
DB 1052 caaaaaaggaagaagaagcagaagtaacagctctccacaagaagcagcctcccgcgcgca 1111  
QY 1200 gggaggttctgagtcgaagtgaaagttggaacacacatgcaagaagaagcagcacaaca 1259  
DB 1112 gggaggttctgagtcgaagtgaaagttggaacacacatgcaagaagaagcagcacaaca 1171  
QY 1260 aagaagctgagcagaagcgaatgcaagtcgaagaacatgctgtagatctctgtttcaaatg 1319  
DB 1172 aagaagctgagcagaagcgaatgcaagtcgaagaacatgctgtagatctctgtttcaaatg 1231  
QY 1320 cgcagcgcagcagccacccaacccagcactcaagtcgaagagagaagacccaataagaac 1379  
DB 1232 cgcagcgcagcagccacccaacccagcactcaagtcgaagagagaagacccaataagaac 1291  
QY 1380 cagggagctggaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1439  
DB 1292 cagggagctggaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1351  
QY 1440 gtaataaagaagatgaagtcaggaatgcttatctaagatcaagcagcgtgctgtcgaa 1499  
DB 1352 gtaataaagaagatgaagtcaggaatgcttatctaagatcaagcagcgtgctgtcgaa 1411  
QY 1500 ttcttcccatggtgcccagagctgcagcagcgtgtagaggttagtcaagagacatacacca 1559  
DB 1412 ttcttcccatggtgcccagagctgcagcagcgtgtagaggttagtcaagagacatacacca 1471  
QY 1560 aagatttcaacagagcagctccgaatctgccaagggccaagcagcagcagcagcagc 1619  
DB 1472 aagatttcaacagagcagctccgaatctgccaagggccaagcagcagcagcagcagc 1531  
QY 1620 gagaggttgttgaatgagggcagcagcagcagcagcagcagcagcagcagcagcagc 1679  
DB 1532 gagaggttgttgaatgagggcagcagcagcagcagcagcagcagcagcagcagcagc 1591  
QY 1680 ctcaagccagctaccacatgagcctctcaagagacccctctgagaacatgactactt 1739

Db 1592 cttcagcagcagctacccacagctacccctcagcagccctctgacacgtgactatctt 1651  
 Qy 1740 ccagagtcagaggttccaagttgtaatacaaaagacttccccaaaaaacaagaacgaat 1799  
 Db 1652 ccagagtcagaggttccaagttgtaatacaaaagacttccccaaaaaacaagaacgaat 1711  
 Qy 1800 ttgtatctcttcaatctgctcctcctcagccacccctgtatccagcaatgtatccgcaag 1859  
 Db 1712 ttgtatctcttcaatctgctcctcctcagccacccctgtatccagcaatgtatccgcaag 1771  
 Qy 1860 atgtgagagtcctgccaatgataatgtgctgcgtgacaacttaagtgtgctgtgagttg 1919  
 Db 1772 atgtgagagtcctgccaatgataatgtgctgcgtgacaacttaagtgtgctgtgagttg 1831  
 Qy 1920 accaacaagaatagagagtgccaagaagaagaagaagaagaagaagaagaagaagaagaaga 1979  
 Db 1832 accaacaagaatagagagtgccaagaagaagaagaagaagaagaagaagaagaagaagaaga 1891  
 Qy 1980 gctgagaccttctgccaatgataatgataatgataatgataatgataatgataatgataatg 2039  
 Db 1892 gctgagaccttctgccaatgataatgataatgataatgataatgataatgataatgataatg 1951  
 Qy 2040 aaactgcttgaataatggaatcttctgatacctccagtgaggcgagagacagcggtgagtt 2099  
 Db 1952 aaactgcttgaataatggaatcttctgatacctccagtgaggcgagagacagcggtgagtt 2011  
 Qy 2100 aaagagatgtggcagcag 2159  
 Db 2012 aaagagatgtggcagcag 2071  
 Qy 2160 tggagactgtgctggggttctgt 2219  
 Db 2072 tggagactgtgctggggttctgt 2131  
 Qy 2220 gcaagctgctgtggggaagaagagagagagagagagagagagagagagagagagagag 2279  
 Db 2132 gcaagctgctgtggggaagaagagagagagagagagagagagagagagagagagagag 2191  
 Qy 2280 atccacacccctggcagccctggttggcttcttcttcttcttcttcttcttcttcttcttct 2339  
 Db 2132 atccacacccctggcagccctggttggcttcttcttcttcttcttcttcttcttcttctt 2251  
 Qy 2340 ggcagagaccccttccaagctgctcaactcagacacacatgggacaacacccctggacagca 2399  
 Db 2252 ggcagagaccccttccaagctgctcaactcagacacacatgggacaacacccctggacagca 2311  
 Qy 2400 tgcag 2459  
 Db 2312 tgcag 2371  
 Qy 2460 tcttactcagcgtgagcccaacttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2519  
 Db 2372 tcttactcagcgtgagcccaacttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2431  
 Qy 2520 ccac 2579  
 Db 2432 ccac 2491  
 Qy 2580 tcttctgagtgatgatacactataataaatttcatcttgagaatttccaattgtatcta 2639  
 Db 2492 tcttctgagtgatgatacactataataaatttcatcttgagaatttccaattgtatcta 2551  
 Qy 2640 gttaaatagcagagtttggaactgtgtgagagagagagagagagagagagagagagagagag 2699  
 Db 2552 gttaaatagcagagtttggaactgtgtgagagagagagagagagagagagagagagagagag 2611  
 Qy 2700 aagatcatatcatatgt 2759  
 Db 2612 aagatcatatcatatgt 2671  
 Qy 2760 ttctgagtgatgcaaatgt 2819

Db 2672 ttctgagtgatgcaaatgt 2731  
 Qy 2820 acctggaataaattctccacagagttgataaattgggcaatttaggaatttaactttag 2879  
 Db 2732 acctggaataaattctccacagagttgataaattgggcaatttaggaatttaactttag 2791  
 Qy 2880 atgatcatattgttccattttcatatttcatatttcatatttcatatttcatatttcatatttcatattt 2939  
 Db 2792 atgatcatattgttccattttcatatttcatatttcatatttcatatttcatatttcatatttcatattt 2851  
 Qy 2940 atgaacttgatctctgttttaagaattatlaaaaaaatgtgtatctatataatgtgc 2999  
 Db 2852 atgaacttgatctctgttttaagaattatlaaaaaaatgtgtatctatataatgtgc 2911  
 Qy 3000 tcttggagagcttagcttctacacactacagagatagatctcagatgtatgtatccataaac 3059  
 Db 2912 tcttggagagcttagcttctacacactacagagatagatctcagatgtatgtatccataaac 2971  
 Qy 3060 ctgcagagtgatcttccacagagtgctcgatactgtlaattacatctccatagggcgaa 3119  
 Db 2972 ctgcagagtgatcttccacagagtgctcgatactgtlaattacatctccatagggcgaa 3031  
 Qy 3120 agaatgacctaagcttctgtataacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3179  
 Db 3032 agaatgacctaagcttctgtataacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3091  
 Qy 3180 gaagtgtgtcagcttaggctgt 3239  
 Db 3092 gaagtgtgtcagcttaggctgt 3151  
 Qy 3240 ttaagtaacgcttccatcttcatattgttttaacgctgtgaatttcttcccataggaatgata 3299  
 Db 3152 ttaagtaacgcttccatcttcatattgttttaacgctgtgaatttcttcccataggaatgata 3211  
 Qy 3300 aaacttaagtgtgtgtcatcatatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 3348  
 Db 3212 aaacttaagtgtgtgtcatcatatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 3260

RESULT 2  
 293323  
 ID 293323 standard; cDNA: 3217 BP.  
 AC 293323;  
 XX  
 DT 04-JUL-2000 (first entry)  
 XX  
 DE Human staufen cDNA.  
 XX  
 KW Staufen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;  
 KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;  
 KW acquired immune deficiency syndrome; protease; human; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 363..1853  
 FT /tag=a  
 FT /product= Staufen  
 XX  
 CA2238656-A1.  
 PN  
 XX  
 PD 22-NOV-1999.  
 XX  
 PF 22-MAY-1998; 98CA-2238656.  
 XX  
 PR 22-MAY-1998; 98CA-2238656.  
 XX  
 PA (UIMO-) UNIV MONTREAL.  
 XX  
 PI Duchaine T, Wickham L, Cohen EA, Luo M, Desgroselliers L;  
 PI Moulard AJ;  
 XX

DR WPI: 2000-246924/22.  
P-PSDB: Y83108.  
PT Isolated mammalian or Caenorhabditis elegans staufen proteins useful  
PT for treating retroviral infections especially human immunodeficiency  
PT virus infections (HIV)  
PS Claim 4; Figure 1a; 96pp; English.  
XX  
CC Staufen is a RNA binding protein which interacts with double stranded  
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA  
CC transport and localization. Mammalian SP contain multiple ds  
CC RNA-binding domains. Chimeric proteins comprising portions of  
CC staufen with proteins with RNase or protease activity can be used to  
CC treat virus infections. The RNase or protease activity of the fusion  
CC protein prevents proper maturation of the virus. Mammalian staufen  
CC proteins seem to recognise double stranded RNA structure rather than  
CC any sequence specific position. 2 bacterially-expressed fusion  
CC proteins used in an RNA-binding assay (his/hsp (human SP) and  
CC MBP/msp (murine SP)) both proteins strongly bound double stranded  
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and  
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the  
CC treatment of RNA virus infections, especially those caused by  
CC retroviruses, in particular human immunodeficiency virus. This  
CC sequence is an alternatively spliced sequence to the one given in  
CC Z93322.  
XX  
SQ Sequence 3217.BP: 889 A: 796 C: 742 G: 790 T: 0 other:  
  
Query Match 91.9%; Score 3076; DB 21; Length 3217;  
Best Local Similarity 96.1%; Pred.No. 0;  
Matches 3217; Conservative 0; Mismatches 0; Indels 131; Gaps 1;  
  
QY 1 actctctcgagcggtgctgagcgctgagcgtcttcagcgtttgctgagcggtgctgagcgt 60  
Db 1 actctctcgagcggtgctgagcgctgagcgtcttcagcgtttgctgagcggtgctgagcgt 60  
QY 61 ctctctcgagcggtgctgagcgctgagcgtcttcagcgtttgctgagcggtgctgagcgt 120  
Db 61 ctctctcgagcggtgctgagcgctgagcgtcttcagcgtttgctgagcggtgctgagcgt 120  
QY 121 ctccagcgcaactcgctcttcctcctcctcctcctcctcctcctcctcctcctcctcct 180  
Db 121 ctccagcgcaactcgctcttcctcctcctcctcctcctcctcctcctcctcctcctcct 180  
QY 181 ctctctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 240  
Db 181 ctctctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 240  
QY 241 agcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 300  
Db 241 agcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 300  
QY 301 gtgactcgtcactcgtcaactccaccctcccaagtcagcgatttcccaattcagcctc 360  
Db 247 ----- 246  
QY 361 ccgataagctgagattaaagatttaaacacttaacactcctcgaactgaacaaagac 420  
Db 247 -----cagagttatttaacacttaacactcctcgaactgaacaaagac 289  
QY 421 aacattgtctcgtgaagcctctttttaaaaaaagaagcaaacctactgtagaact 480  
Db 290 aacattgtctcgtgaagcctctttttaaaaaaagaagcaaacctactgtagaact 349  
QY 481 aaatgacactgctgaacttgaaaaaaacaaatgataagcctgtgaccccttactc 540  
Db 350 aaatgacactgctgaacttgaaaaaaacaaatgataagcctgtgaccccttactc 409  
QY 541 tcgagatcagctcaactataactacaatgagagaggtgtcttaccgccgaggtactc 600  
Db 410 tcgagatcagctcaactataactacaatgagagaggtgtcttaccgccgaggtactc 469

QY 601 ttaccattccagttccacccttacttatacaagtggaacttctgtggagagacagca 660  
Db 470 ttaccattccagttccacccttacttatacaagtggaacttctgtggagagacagca 529  
QY 661 attaatgcaaaagaaagacaaagacagcgctgcgaacacagatgctgccaagcgct 720  
Db 530 attaatgcaaaagaaagacaaagacagcgctgcgaacacagatgctgccaagcgct 589  
QY 721 gaggatctcgaagaatgagccctgcgagagagctggaagtgatgtaaaagaatccga 780  
Db 590 gaggatctcgaagaatgagccctgcgagagagctggaagtgatgtaaaagaatccga 649  
QY 781 agaaataatcccaataatcgaataatgcaaatgtttgagattgcaacttaacgaa 840  
Db 650 agaaataatcccaataatcgaataatgcaaatgtttgagattgcaacttaacgaa 709  
QY 841 ctgtcgttgaaattcgaagtgccgggagagtgccacccacatgaaagaacttctg 900  
Db 710 ctgtcgttgaaattcgaagtgccgggagagtgccacccacatgaaagaacttctg 769  
QY 901 gaccaggttcggttgaggttggttggttggttggttggttggttggttggttggttg 960  
Db 770 gaccaggttcggttgaggttggttggttggttggttggttggttggttggttggttg 829  
QY 961 aaagaaatcccgccatagctgttcttgagagctggaagaagttaacgcccctgctgc 1020  
Db 830 aaagaaatcccgccatagctgttcttgagagctggaagaagttaacgcccctgctgc 889  
QY 1021 agtltgaacgaataaagcctatgaatcaaaagaaacaaacccaatgataagccacagac 1080  
Db 890 agtltgaacgaataaagcctatgaatcaaaagaaacaaacccaatgataagccacagac 949  
QY 1081 aagcccaaatatggtccagggagatcaatccgattgacgactggtgcccagatccagcagc 1140  
Db 950 aagcccaaatatggtccagggagatcaatccgattgacgactggtgcccagatccagcagc 1009  
QY 1141 aaaaaagagaagagagcagagtaacagctcctcaagaagcctcccgcgcgacag 1200  
Db 1010 aaaaaagagaagagagcagagtaacagctcctcaagaagcctcccgcgcgacag 1069  
QY 1201 ggaattgtgatacagtgaaagtgtaaacacacatgcagaagaagcagcgacacaa 1260  
Db 1070 ggaattgtgatacagtgaaagtgtaaacacacatgcagaagaagcagcgacacaa 1129  
QY 1261 gaaagtgtccaaagcgaatgtcagccgaagaacatgtgtgaagtccttggttcaaaagtc 1320  
Db 1130 gaaagtgtccaaagcgaatgtcagccgaagaacatgtgtgaagtccttggttcaaaagtc 1189  
QY 1321 gcaagcgagcccaacaaacccgcaatcaagtcagaagagaagacacccaataaagaacc 1380  
Db 1190 gcaagcgagcccaacaaacccgcaatcaagtcagaagagaagacacccaataaagaacc 1249  
QY 1381 agggagtgaaagaaagtaacctttttgaacctggtctctgggagtgaaatggagctag 1440  
Db 1250 agggagtgaaagaaagtaacctttttgaacctggtctctgggagtgaaatggagctag 1309  
QY 1441 taataaagagatgagttcagaatgacctatcctatgaatgaagcagcgtcgtcgtgagat 1500  
Db 1310 taataaagagatgagttcagaatgacctatcctatgaatgaagcagcgtcgtcgtgagat 1369  
QY 1501 tcttcccatgtgcccggagtcgcccgaagctgtlagaggtttagtcaagagacatcacacaa 1560  
Db 1370 tcttcccatgtgcccggagtcgcccgaagctgtlagaggtttagtcaagagacatcacacaa 1429  
QY 1561 agatttaccagagcgtccgaatcctgtcagaagcgaagtgtaactgcataatagcccg 1620  
Db 1430 agatttaccagagcgtccgaatcctgtcagaagcgaagtgtaactgcataatagcccg 1489  
QY 1621 agagttgttataatggggcgaactcgcgcacagccgaagcgaattttaaagaataaacttc 1680  
Db 1490 agagttgttataatggggcgaactcgcgcacagccgaagcgaattttaaagaataaacttc 1549



Db 507 cagcaatttaatgccaagaaagacagctgcgaaacacgatgtgtgtccaa 566

.....

[illegible][illegible]

RESULT	4	
ID	293322	Standard; cDNA: 3506 BP.
XX	293322;	
XX	AC	
XX	04-JUL-2000	(first entry)
DT		
XX	Human stauften cDNA.	
DE		
XX	Staufen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;	
KW	HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;	
KW	acquired immune deficiency syndrome; protease; human; ss.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	409..2142
FT		/tag=a
FT		/product= Staufen
XX		
PN	CA2238656-A1.	
XX		
PD	22-NOV-1999.	
XX		
PF	22-MAY-1998;	98CA-2238656.
XX		
PR	22-MAY-1998;	98CA-2238656.
XX		
PA	(UYMO-) UNIV MONTREAL.	
XX		
PI	Duchaine T, Wickham L, Cohen EA, Luo M, Desrosiers L;	
PI	Mouland AJ;	
XX		
DR	WPI: 2000-246924/22.	
DR	P-PSDB; Y83023.	
XX		
PT	Isolated mammalian or Caenorhabditis elegans stauften proteins useful	
PT	for treating retroviral infections especially human immunodeficiency	
XX	virus infections (HIV)	
XX		

PS Claim 4; Figure 1a; 96pp; English.  
XX  
CC Staufen is a RNA binding protein which interacts with double stranded  
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA  
CC transport and localization. Mammalian SP contain multiple ds  
CC RNA-binding domains. Chimeric proteins comprising portions of  
CC staufen with proteins with RNase or protease activity can be used to  
CC treat virus infections. The RNase or protease activity of the fusion  
CC protein prevents proper maturation of the virus. Mammalian staufen  
CC proteins seem to recognise double stranded RNA structure rather than  
CC any sequence specific position. 2 bacterially-expressed fusion  
CC proteins used in an RNA-binding assay (his/hsp (human SP) and  
CC MBP/msp (murine SP)) both proteins strongly bound double stranded  
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and  
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the  
CC treatment of RNA virus infections, especially those caused by  
CC retroviruses, in particular human immunodeficiency virus.  
XX  
SQ Sequence 3506 BP; 965 A; 875 C; 790 G; 876 T; 0 other;

Query Match 88.5%; Score 2961.6; DB 21; Length 3506;  
Best Local Similarity 92.0%; Pred. No. 0;  
Matches 3224; Conservative 0; Mismatches 124; Indels 158; Gaps 3;

OY 1 actctctccggcgctgctgagcgctctcaagcgtttgcgcgcgctgcgct 60  
DB 1 attctctcgcggtcgctgagcgctctcaagcgtttgcgcgcgctgcgct 60  
OY 61 ctctctcgctcgctctctcttgaacgcctcccccgcgcgcgcgcgcctc 120  
DB 61 ctctctcgctcgctctctcttgaacgcctcccccgcgcgcgcgcgcctc 120  
OY 121 ctccacggcactccgctctctctctctctctctctctctctctctctct 180  
DB 121 ctccacggcactccgctctctctctctctctctctctctctctctctct 180  
OY 181 ctctctcccccctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 240  
DB 181 ctctctcccccctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 240  
OY 241 agcagccag----- 249  
DB 241 agcagccaggtttatttaaccacttaaccttcagaactgaacaagaacatgttcc 300  
OY 250 -----cagcagccaggtgtgttctct----- 275  
DB 301 tggaaagcctcttttaaaaaagtagaacttagacttcaatgaactgaactcgt 360  
OY 276 -----gtcgccagcgtgagtg 293  
DB 361 cactgaagcgttttaacctgttcaactgttcaactgttgaagtagaactgtcctaagt 420  
OY 294 cagtgctgtatctcgctcaactgaacctcaacctccaggtcagcgatlttccact 353  
DB 421 caagtgcaagttcagaaccactgtctctctcagggagcaatactgaacaagaac 480  
OY 354 cagcgtcccgataagctgagattacagagttataaaccact----- 395  
DB 481 cagtcctctctcacaagcctttagtgattctcttcaactacagactctctgcctct 540  
OY 396 -----taactctcagaactgaacaagaacaactgttctctggaagcctt 442  
DB 541 gaaaaatgagtagaaccacttaaacacttgcttacccttgcatctataatcacc 600  
OY 443 cttttaaaaaagaagactaaaccctactgtagaactaaatgactgtgataaact 502  
DB 601 agtgcagctgcgaagaactaaaccctactgtagaactaaatgactgtgataaact 660  
OY 503 ggaaaaaaacaatgtataagcgttgaaccttaactcagagtagtccactataac 562  
DB 661 ggaaaaaaacaatgtataagcgttgaaccttaactcagagtagtccactataac 720

OY 563 tacacaatgagagagtgcttatcccccaggtacttlaaccattccagttccacct 622  
DB 721 tacaacatgagagagtgcttatcccccaggtacttlaaccattccagttccacct 780  
OY 623 ttaacttcaagtggaactctctgtggaagacagcaattaatgtccaaaggaagaca 682  
DB 781 ttacttlaagtggaactctctgtggaagacagcaattaatgtccaaaggaagaca 840  
OY 683 agacagctgcgcaaacacagatctctctccaaagcgttgaagatctccagaatgagcc 742  
DB 841 agacagctgcgcaaacacagatctctctccaaagcgttgaagatctccagaatgagcc 900  
OY 743 ctgcgaagagctgtgaggtgaatgaagagatccgaagaagaataatcctaataact 802  
DB 901 ctgcgaagagctgtgaggtgaatgaagagatccgaagaagaataatcctaataact 960  
OY 803 gaataagtcaggtgttgagatgtcacttaacgggaacttgctctgtaattcgaagt 862  
DB 961 gaataagtcaggtgttgagatgtcacttaacgggaacttgctctgtaattcgaagt 1020  
OY 863 gcccgagagagtgcccccacatgaagaactttgtgaccaaagtttcggttggag 922  
DB 1021 gcccgagagagtgcccccacatgaagaactttgtgaccaaagtttcggttggag 1080  
OY 923 ttgtgagggaggtgaagggaagaagcaagatltcaagaagaataatgcccgaact 982  
DB 1081 ttgtgagggaggtgaagggaagaagcaagatltcaagaagaataatgcccgaact 1140  
OY 983 gttcttgaagagctgaagaagttaccgccccgctgtcagttgaaacggtaaagcttga 1042  
DB 1141 gttcttgaagagctgaagaagttaccgccccgctgtcagttgaaacggtaaagcttga 1200  
OY 1043 atcaaaagaaacaacaacccaatgaatcaagccacagacagccagaatatgtgccaggg 1102  
DB 1201 atcaaaagaaacaacaacccaatgaatcaagccacagacagccagaatatgtgccaggg 1260  
OY 1103 atcaatccgattagccagctgtgccagatccagcagcgcaaaaaagagagagccagag 1162  
DB 1261 atcaatccgattagccagctgtgccagatccagcagcgcaaaaaagagagagccagag 1320  
OY 1163 tacaagctcttcaagagagcagagcttccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1222  
DB 1321 tacaagctcttcaagagagcagagcttccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1380  
OY 1223 gttgaaacccacacatgcggaagaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1282  
DB 1381 gttgaaacccacacatgcggaagaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1440  
OY 1283 gccgaagacatgtctgagatctctgttcaagttcccgacgcgcgcgcgcgcgcgcgcgcgc 1342  
DB 1441 gccgaagacatgtctgagatctctgttcaagttcccgacgcgcgcgcgcgcgcgcgcgcgc 1500  
OY 1343 gcaatcaagtcagagagagagacaccataaagaagaacccagggatgtgaaagaataaac 1402  
DB 1501 gcaatcaagtcagagagagagacaccataaagaagaacccagggatgtgaaagaataaac 1560  
OY 1403 tttttgaacctgtctgtgggataaaatgagactagttaaaagagagatgaagtccag 1462  
DB 1561 tttttgaacctgtctgtgggataaaatgagactagttaaaagagagatgaagtccag 1620  
OY 1463 atgcttatcagtcacagcagctgtcgtgtgaattcttcccatgtgtgcgcaggttc 1522  
DB 1621 atgcttatcagtcacagcagctgtcgtgtgaattcttcccatgtgtgtgcgcaggttc 1680  
OY 1523 gccacagctgtagagtagtcaaggacatcacacaaagatttaccagggcagctccg 1582  
DB 1681 gccacagctgtagagtagtcaaggacatcacacaaagatttaccagggcagctccg 1740  
OY 1583 aatctgcgaagccaggttaactgcatatgataagccagaggtgtgtatgtggggcacc 1642  
DB 1741 aatctgcgaagccaggttaactgcatatgataagccagaggtgtgtatgtggggcacc 1800  
OY 1643 tgcgccaagccgagacatttlaagaaataacatctcttcaagcgcaagctaccatgtga 1702



Db	1801	tcgcccacagcagagacacattttaaagaatacaacctctccagcgccacgtaacccacgga	1660
QY	1703	cccttcacgagaccctctcgacacacgagactactcttcacagatgccaggaattccagtt	1762
Db	1861	cccttcacgagaccctctcgacacacgagactactcttcacagatgccaggaattccagtt	1920
QY	1763	gaatacaagaactctcccaaaaaacaagaagaacttgatctctatcaattgtccc	1822
Db	1921	gaatacaagaactctcccaaaaaacaagaagaacttgatctctatcaattgtccc	1980
QY	1823	ttctagccacctctgatacagccatgtaacggcagagatgtgagatctctgcatagtatg	1882
Db	1961	ttctagccacctctgatacagccatgtaacggcagagatgtgagatctctgcatagtatg	2040
QY	1883	gctgcgctggaacatcttaagaagtgtgcgtctgttgagaccacaagaatgacagatbtcca	1942
Db	2041	gctgcgctggaacatcttaagaagtgtgcgtctgttgagaccacaagaatgacagatbtcca	2100
QY	1943	agaacagagaacacggaacccaatgtctgtgtgtgtgagagtgctgaaccttcttcgacatgaa	2002
Db	2101	agaacagagaacacggaacccaatgtctgtgtgtgtgagagtgctgaaccttcttcgacatgaa	2160
QY	2003	ccattataaaaatcccaacataatacttgaaataactgcgtcttgaaaaatttgaaat	2062
Db	2161	ccattataaaaatcccaacataatacttgaaataactgcgtcttgaaaaatttgaaat	2220
QY	2063	ttctcgataacctccagctgtgagcgacgaagaacacggctgggttaagaatatgtggacagacagag	2122
Db	2221	ttctcgataacctccagctgtgagcgacgaagaacacggctgggttaagaatatgtggacagacag	2280
QY	2123	aagaacaaacagaacaacaaaggaagcgagctgtgtggccgctgtgaactgtgcgagaggtttgtgt	2182
Db	2281	aagaacaaacagaacaacaaaggaagcgagctgtgtggccgctgtgaactgtgcgagaggtttgtgt	2340
QY	2183	gagtggccaactcgtgtgacctgtgcggttccctaagcaatagcagctgtccgtgtgggagaagag	2242
Db	2341	gagtggccaactcgtgtgacctgtgcggttccctaagcaatagcagctgtccgtgtgggagaagag	2400
QY	2243	gctgcgccagacagctgtgttctcccggaacaacagacatccacaacccctgggacactccgt	2302
Db	2401	gctgcgccagacagctgtgttctcccggaacaacagacatccacaacccctgggacactccgt	2460
QY	2303	gtttggtctcttttcttccctctgtgtgaagaagaagaacgacagacccctctcaagctcgt	2362
Db	2461	gtttggtctcttttcttccctctgtgtgaagaagaagaacgacagacccctctcaagctcgt	2520
QY	2363	ctcacacagacaacattvggagacaaacccctvgagacagccatgccaagagagagcccttbgacg	2422
Db	2521	ctcacacagacaacattvggagacaaacccctvgagacagccatgccaagagagagcccttbgacg	2580
QY	2423	gcccagagacctaanaagacaacagagaagaataccaatgtcttctactaagtgatgaacccaatt	2482
Db	2581	gcccagagacctaanaagacaacagagaagaataccaatgtcttctactaagtgatgaacccaatt	2640
QY	2483	ttctcagtggtgcacaagggcccacacacactcctctgcagtaaccacacataaccactgtcttct	2542
Db	2641	ttctcagtggtgcacaagggcccacacacactcctctgcagtaaccacacataaccactgtcttct	2700
QY	2543	cttcccaacagtatctgtatcttctagtttcaatatttctttagatgataatgacatat	2602
Db	2701	cttcccaacagtatctgtatcttctagtttcaatatttctttagatgataatgacatat	2760
QY	2603	ataaatttcaatttgagaatttctcaattgtatctagtttaataatgacagattvggaac	2662
Db	2761	ataaatttcaatttgagaatttctcaattgtatctagtttaataatgacagattvggaac	2820
QY	2663	ttgtctcgagactgacttataacaaataacaaacggaacaaagacatatacagatgtgatatg	2722
Db	2821	ttgtctcgagactgacttataacaaataacaaacggaacaaagacatatacagatgtgatatg	2880
QY	2723	gttagacaattttaaatttcaatttgataccacagagacggtttcagatgataaattgtgtc	2782

Db	2881	gttagacattttattttatcttcaatgactaacccagagacagtttcagtgatgcaattgtgtgc	2940
Qy	2783	cctctgtgtcagctgtaaaacagctcctggaattccaaaaaacctgtaaaagctccccaagt	2842
Db	2941	ccctctgtgtcagctgtaaaacagctcctggaattccaaaaaacctgtaaaagctccccaagt	3000
Qy	2843	tgtataaattgagacaatttaggaattttaactttagatgatcatcttggttccatttta	2902
Db	3001	tgtataaattgagacaatttaggaattttaactttagatgatcatcttggttccatttta	3060
Qy	2903	tttaattttatttttggtaatgcaaacagagcttaaatgaactttgcttcgttttaa	2962
Db	3061	tttaattttatttttggtaatgcaaacagagcttaaatgaactttgcttcgttttaa	3120
Qy	2963	agattatataaaaacattgtgtatactatacatatgctcttgaagacttaagcttcaacta	3022
Db	3121	agattatataaaaacattgtgtatactatacatatgctcttgaagacttaagcttcaacta	3180
Qy	3023	caactacagagatagatctccatgtagtaccataataaacctgcagagtgatttccagagtg	3082
Db	3181	caactacagagatagatctccatgtagtaccataataaacctgcagagtgatttccagagtg	3240
Qy	3083	ctcgatactgttaattatcatcatcccatatgaggtcgtgaagaagaatgacctagtttcgtata	3142
Db	3241	ctcgatactgttaattatcatcatcccatatgaggtcgtgaagaagaatgacctagtttcgtata	3300
Qy	3143	cagcgtggtgttccttgaatgtgtgttaactgtacacagagagtgctgtgacctgaagctctg	3202
Db	3301	cagcgtggtgttccttgaatgtgtgttaactgtacacagagagtgctgtgacctgaagctctg	3360
Qy	3203	cggtgtgtccgtatgagaaaacctgtgagccctgcgagtgtaagttaactgtcttccatcatctg	3262
Db	3361	cggtgtgtccgtatgagaaaacctgtgagccctgcgagtgtaagttaactgtcttccatcatctg	3420
Qy	3263	tttaagctgtgaattttcttccccaatggaatgtgaatgaatgaatgaatgtgtgtcatcaat	3322
Db	3421	tttaagctgtgaattttcttccccaatggaatgtgaatgaatgaatgaatgtgtgtcatcaat	3480
Qy	3323	aaatggtatatactataaaaaaaa 3348	
Db	3481	aaatggtatatactataaaaaaaa 3506	
RESULT 5			
ID	Z93324	Z93324 standard; cDNA; 3142 BP.	
XX	XX	Z93324;	
AC	XX		
XX	XX		
DT	XX	04-JUL-2000 (first entry)	
XX	XX		
DE	XX	Human stauflen cDNA.	
XX	XX		
KM	XX	Staufen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;	
XX	XX	HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;	
KM	XX	acquired immune deficiency syndrome; protease; human; ss.	
XX	XX		
OS	XX	Homo sapiens.	
XX	XX		
FN	XX	Key	
FT	XX	CDS	
FT	XX	Location/Qualifiers	
FT	XX	288..1193	
FT	XX	/*tag= a	
FT	XX	/product= Staufen protein	
XX	XX	CA2238656-A1.	
XX	XX		
PD	XX	22-NOV-1999.	
XX	XX		
PE	XX	22-MAY-1998; 98CA-2238656.	
XX	XX		
PR	XX	22-MAY-1998; 98CA-2238656.	
XX	XX		
PA	XX	(UYMO-) UNIV MONTREAL.	

XX Duchaine T, Wickham L, Cohen EA, Luo M, Deesrosellers L;  
PI Moulard AJ;  
XX  
DR WPI: 2000-246924/22.  
DR P-PSDB: Y83108.

XX Isolated mammalian or Caenorhabditis elegans staufen proteins useful  
XX for treating retroviral infections especially human immunodeficiency  
XX virus infections (HIV)

PS Claim 4: Figure 1b; 96pp; English.

XX Staufen is a RNA binding protein which interacts with double stranded  
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA  
CC transport and localization. Mammalian SP contain multiple ds  
CC RNA-binding domains. Chimeric proteins comprising portions of  
CC staufen with proteins with RNase or protease activity can be used to  
CC treat virus infections. The RNase or protease activity of the fusion  
CC protein prevents proper maturation of the virus. Mammalian staufen  
CC proteins seem to recognise double stranded RNA structure rather than  
CC any sequence specific position. 2 bacterially-expressed fusion  
CC proteins used in an RNA-binding assay (his/NSP (human SP) and  
CC MBP/msp (murine SP)) both proteins strongly bound double stranded  
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and  
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the  
CC treatment of RNA virus infections, especially those caused by  
CC retroviruses, in particular human immunodeficiency virus. This  
CC sequence is an alternatively spliced sequence to the one given in  
CC Y93322.

XX Sequence 3142 BP; 862 A; 778 C; 733 G; 769 T; 0 other;

Query Match 87.4%; Score 2926; DB 21; Length 3142;  
Best Local Similarity 93.8%; Pred. No. 0;

Matches 3142; Conservative 0; Mismatches 0; Indels 206; Gaps 1;

QY 1 attctctgcggcgctgctgagcgctctcagcgatttgcgcgcgctcgcgct 60  
DB 1 attctctgcggcgctgagcgctctcagcgatttgcgcgcgctcgcgct 60  
QY 61 ctctctgcgctccgctctcttgacgcgctcccccgcgcgcgcgcgcgctc 120  
DB 61 ctctctgcgctccgctctcttgacgcgctcccccgcgcgcgcgcgcgctc 120  
QY 121 ctccacgagcactcgctctctccctctgctcccttctcccttctccct 180  
DB 121 ctccacgagcactcgctctctccctctgctcccttctcccttctccct 180  
QY 181 ctctctccctccctcgccgacgcgcgcgcgcgcgcgcgcgcgcgcgcgc 240  
DB 181 ctctctccctccctcgccgacgcgcgcgcgcgcgcgcgcgcgcgcgcgc 240  
QY 241 agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 300  
DB 241 agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 300  
QY 301 gtgacatcggtcactgcaactccacgtccaggtcaggaatttccactcagc 360  
DB 301 gtgacatcggtcactgcaactccacgtccaggtcaggaatttccactcagc 360  
QY 361 ccgataagctgagattacagagttataaacacttaaccttcgaacttgaaagac 420  
DB 361 ccgataagctgagattacagagttataaacacttaaccttcgaacttgaaagac 420  
QY 421 aacattgtctcctggaagcctcttttaaaaagaagacataacccactactaga 480  
DB 421 aacattgtctcctggaagcctcttttaaaaagaagacataacccactactaga 480  
QY 481 aaatgcactgtgacatgaaacttgaaaaaaccaatgataagcctgttgacccttact 540  
DB 481 aaatgcactgtgacatgaaacttgaaaaaaccaatgataagcctgttgacccttact 540  
QY 541 tcgagatgcagttccacctataactacacatgagagagtgcttatacccccaggtact 600  
DB 541 tcgagatgcagttccacctataactacacatgagagagtgcttatacccccaggtact 600  
QY 601 ttaccatttcagttccacctataactataacttgaaacttctctgtgagagacga 660  
DB 601 ttaccatttcagttccacctataactataacttgaaacttctctgtgagagacga 660  
QY 661 atttaatgcaagagaaagacaaagcagcttgcaaacccagctgctgcgaagcgt 720  
DB 661 atttaatgcaagagaaagacaaagcagcttgcaaacccagctgctgcgaagcgt 720  
QY 721 gaggatcccgcaagatgagcccttcgcaagagagcttgagagtgaaagagaaatccga 780  
DB 721 gaggatcccgcaagatgagcccttcgcaagagagcttgagagtgaaagagaaatccga 780  
QY 781 agagaagaaatcctaataaactgaaataagtgaaagtggttgagattgcaacttaacgaa 840  
DB 781 agagaagaaatcctaataaactgaaataagtgaaagtggttgagattgcaacttaacgaa 840  
QY 841 ctgctctggaatttcgagagtgcccggaagtgccacccacatgaaagacttctgt 900  
DB 841 ctgctctggaatttcgagagtgcccggaagtgccacccacatgaaagacttctgt 900  
QY 901 gaccaaagcttcggttgaggagcttgctgaggagagtgaaagagaaagcaagaatcttc 960  
DB 901 gaccaaagcttcggttgaggagcttgctgaggagagtgaaagagaaagcaagaatcttc 960  
QY 961 aaagaagaaatgcgcacataagctgtctcttgagagagcttgaaagagttaccgcccctgc 1020  
DB 961 aaagaagaaatgcgcacataagctgtctcttgagagagcttgaaagagttaccgcccctgc 1020  
QY 1021 agttgaagagtaagagcttagatataaaagaaacaaacccatagtcgaagccacagac 1080  
DB 1021 agttgaagagtaagagcttagatataaaagaaacaaacccatagtcgaagccacagac 1080  
QY 815 agttgaagagtaagagcttagatataaaagaaacaaacccatagtcgaagccacagac 874  
DB 815 agttgaagagtaagagcttagatataaaagaaacaaacccatagtcgaagccacagac 874  
QY 1081 aagccagagatattgcagagagatcatcgaattgagcagactgagccagatccacagagc 1140  
DB 1081 aagccagagatattgcagagagatcatcgaattgagcagactgagccagatccacagagc 1140  
QY 1141 aaaaaagagagagagagcagagtagacagctctctcacaagagcagagcctcccgcgcgag 1200  
DB 1141 aaaaaagagagagagagcagagtagacagctctctcacaagagcagagcctcccgcgcgag 1200  
QY 1201 ggaatttctgagcagagtgaaagtgtaaacacacttgcaagagaaacgggacacaaacaa 1260  
DB 1201 ggaatttctgagcagagtgaaagtgtaaacacacttgcaagagaaacgggacacaaacaa 1260  
QY 1261 gaaagtgagcagagcagatgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1320  
DB 1261 gaaagtgagcagagcagatgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1320  
QY 1321 gcaagcagagccacacaaacccgcacacaaagtcaagagagaaagacacccataaagaac 1380  
DB 1321 gcaagcagagccacacaaacccgcacacaaagtcaagagagaaagacacccataaagaac 1380  
QY 1381 agggatggaagaaagtaaaccttttgaaccttgctctgagagtgaaatggagactag 1440  
DB 1381 agggatggaagaaagtaaaccttttgaaccttgctctgagagtgaaatggagactag 1440  
QY 1441 taataaagagatgagttcagagatcctatatactataagatcagcagcagctgcgcgagat 1500  
DB 1441 taataaagagatgagttcagagatcctatatactataagatcagcagcagctgcgcgagat 1500  
QY 1501 tctcccatggtgcccgaagtgccgcaagcgtgtagagtttagcaagagacatcacacaa 1560  
DB 1501 tctcccatggtgcccgaagtgccgcaagcgtgtagagtttagcaagagacatcacacaa 1560  
QY 1561 agattttacaaagagagctccgaatcctgccaagagcgaagcaatgcatgataagcccg 1620  
DB 1561 agattttacaaagagagctccgaatcctgccaagagcgaagcaatgcatgataagcccg 1620  
QY 1621 agattttacaaagagagctccgaatcctgccaagagcgaagcaatgcatgataagcccg 1680  
DB 1621 agattttacaaagagagctccgaatcctgccaagagcgaagcaatgcatgataagcccg 1680

QY 1621 agagtgtgtgtatgtgggacacctgcgcccacagccagaccattttaagaataacatctc 1680  
 |||||  
 Db 1415 agagtgtgtgtatgtgggacacctgcgcccacagccagaccattttaagaataacatctc 1474  
 QY 1681 ttcaaggccacgtatcccaatggagacctcaacagacctcttgagcaactgtactatcttc 1740  
 |||||  
 Db 1475 ttcaaggccacgtatcccaatggagacctctcaacagacctcttgagcaactgtactatcttc 1534  
 QY 1741 cagagccagggatcccaagtttgaatacaagaacttccccaaaaaacaagaagaactt 1800  
 |||||  
 Db 1535 cagagtcagggatcccaagtttgaatacaagaacttccccaaaaaacaagaagaactt 1594  
 QY 1801 tgaatctctatcaatctgtcctctcaagccactctgatacagccatgtaactgcgaagga 1860  
 |||||  
 Db 1595 tgaatctctatcaatctgtcctctcaagccactctgatacagccatgtaactgcgaagga 1654  
 QY 1861 tgtgtgagctccatgataatgtgcgtctgtaacatcttaaaagtgtctgtcgaatttga 1920  
 |||||  
 Db 1655 tgtgtgagctccatgataatgtgcgtctgtaacatcttaaaagtgtctgtcgaatttga 1714  
 QY 1921 ccaacaagaatagagagatcccaagaagaaggaacgaactgtctgtgtgtggaggtg 1980  
 |||||  
 Db 1715 ccaacaagaatagagagatcccaagaagaaggaacgaactgtctgtgtgtggaggtg 1774  
 QY 1981 ctgaaaccttctctgagccatgaaacattataaaatcccaacatatactatcgaataactga 2040  
 |||||  
 Db 1775 ctgaaaccttctctgagccatgaaacattataaaatcccaacatatactatcgaataactga 1834  
 QY 2041 aactgctttgaaaaatttggaaatttctgataactcgaatgggcccagagacaggtgtgta 2100  
 |||||  
 Db 1835 aactgctttgaaaaatttggaaatttctgataactcgaatgggcccagagacaggtgtgta 1894  
 QY 2101 aaggaatgtggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2160  
 |||||  
 Db 1895 aaggaatgtggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1954  
 QY 2161 ggaactgtctgtgggttctgt 2220  
 |||||  
 Db 1955 ggaactgtctgtgggttctgt 2014  
 QY 2221 cagctgcctgtgggaggaagaaggtctgccagcagcagcagcagcagcagcagcagcagc 2280  
 |||||  
 Db 2015 cagctgcctgtgggaggaagaaggtctgccagcagcagcagcagcagcagcagcagcagc 2074  
 QY 2281 tccacacccctgtggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2340  
 |||||  
 Db 2075 tccacacccctgtggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2134  
 QY 2341 gcaagaccccttctcaagctgtgtcaactcagacacatctggagacaacccctgtgacagcat 2400  
 |||||  
 Db 2135 gcaagaccccttctcaagctgtgtcaactcagacacatctggagacaacccctgtgacagcat 2194  
 QY 2401 gccagagagaagccttctgacggcccccagagcctaagaagcagagagaaaaatcaatgctt 2460  
 |||||  
 Db 2195 gccagagagaagccttctgacggcccccagagcctaagaagcagagagaaaaatcaatgctt 2254  
 QY 2461 cccaactcagcgtgacccaacttctcagtggtgacagggcccccacccctctgcagctacc 2520  
 |||||  
 Db 2255 cccaactcagcgtgacccaacttctcagtggtgacagggcccccacccctctgcagctacc 2314  
 QY 2521 cacacacatcacactgcttctctctccacagtgactgtatctctagtttcaatttctt 2580  
 |||||  
 Db 2315 cacacacatcacactgcttctctctccacagtgactgtatctctagtttcaatttctt 2374  
 QY 2581 ctttctgtatgataatgacataataaatttcaatttgagaatttctcaattgtatctag 2640  
 |||||  
 Db 2375 ctttctgtatgataatgacataataaatttcaatttgagaatttctcaattgtatctag 2434  
 QY 2641 ttaataagacagatttggaaactgtctgagtgacttataatcaatcaatcaacagcaaa 2700  
 |||||  
 Db 2435 ttaataagacagatttggaaactgtctgagtgacttataatcaatcaatcaacagcaaa 2494  
 QY 2701 agatcatatcagatgtatgtgttagacatttattatcatgtactaacccagcagc 2760

Db 2495 agatcatatcagatgtatgtgttagacatttattatcatgtactaacccagcagc 2554  
 |||||  
 QY 2761 ttcaagtgagcaaatgtgtgcccctgtgtgttcaagcctggaacagcctgtgacttcaaaa 2820  
 |||||  
 Db 2555 ttcaagtgagcaaatgtgtgcccctgtgtgttcaagcctggaacagcctgtgacttcaaaa 2614  
 QY 2821 ccttgataagtcctccacagctgtataaatltgagacaatttgaactttaaactttaga 2880  
 |||||  
 Db 2615 ccttgataagtcctccacagctgtataaatltgagacaatttgaactttaaactttaga 2674  
 QY 2881 tgaatcttggttccatttatttcaatttcaatttcaatttcaatttcaatttcaatttcaatt 2940  
 |||||  
 Db 2675 tgaatcttggttccatttatttcaatttcaatttcaatttcaatttcaatttcaatttcaatt 2734  
 QY 2941 tgaacttgatctctgtttaaagattataaaacattgtgtactataactatgct 3000  
 |||||  
 Db 2735 tgaacttgatctctgtttaaagattataaaacattgtgtactataactatgct 2794  
 QY 3001 cttgagacttgaacttctcaactacacacagatagatctcacaatgtatgtatcataaac 3060  
 |||||  
 Db 2795 cttgagacttgaacttctcaactacacacagatagatctcacaatgtatgtatcataaac 2854  
 QY 3061 tgcagagtgatttccagagtgctcgatactgttaattataatccattagggctgaaaa 3120  
 |||||  
 Db 2855 tgcagagtgatttccagagtgctcgatactgttaattataatccattagggctgaaaa 2914  
 QY 3121 gaatgacctaagcttctgtataacagctgtgtgcttcttgaatgtgtgtgtgtgtgtgtgt 3180  
 |||||  
 Db 2915 gaatgacctaagcttctgtataacagctgtgtgcttcttgaatgtgtgtgtgtgtgtgtgt 2974  
 QY 3181 aagtgtgtgacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3240  
 |||||  
 Db 2975 aagtgtgtgacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3034  
 QY 3241 taagtactgttccatcatcatgttctacgcgtgaaatttctcccatggaatgaa 3300  
 |||||  
 Db 3035 taagtactgttccatcatcatgttctacgcgtgaaatttctcccatggaatgaa 3094  
 QY 3301 aacttaagtgattgtcatcataaataatggtatatacttaaaaaaaaaa 3348  
 |||||  
 Db 3095 aacttaagtgattgtcatcataaataatggtatatacttaaaaaaaaaa 3142  
 |||||  
 RESULT 6  
 Z93328  
 ID Z93328 standard; cDNA; 2859 BP.  
 AC Z93328;  
 XX  
 DT 04-JUL-2000 (first entry)  
 XX  
 DE Mouse stauufen cDNA.  
 XX  
 KW Stauufen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;  
 KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;  
 KW acquired immune deficiency syndrome; protease; mouse; ss.  
 OS Mus musculus.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 325..1788  
 FT /tag= a  
 FT /product= Stauufen protein  
 PN CA2238656-A1.  
 PD 22-NOV-1999.  
 XX  
 PF 22-MAY-1998; 98CA-2238656.  
 XX  
 PR 22-MAY-1998; 98CA-2238656.  
 XX

PA (UYMO-) UNIV MONTREAL.  
 XX Duchaine F, Wickham L, Cohen EA, Luo M, Desrosiers L.  
 PI Moulard AJ.  
 XX WPI: 2000-246924/22.  
 DR P-PSDB: Y83024.  
 XX  
 PI Isolated mammalian or Caenorhabditis elegans staufen proteins useful  
 PI for treating retroviral infections especially human immunodeficiency  
 PI virus infections (HIV)  
 XX  
 XX Claim 4; Figure 1c; 96pp; English.

Staufen is a RNA binding protein which interacts with double stranded  
 CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA  
 CC transport and localization. Mammalian SP contain multiple ds  
 CC RNA-binding domains. Chimeric proteins comprising portions of  
 CC staufen with proteins with RNase or protease activity can be used to  
 CC treat virus infections. The RNase or protease activity of the fusion  
 CC protein prevents proper maturation of the virus. Mammalian staufen  
 CC proteins seem to recognise double stranded RNA structure rather than  
 CC any sequence specific position. 2 bacterially-expressed fusion  
 CC proteins used in an RNA-binding assay (his/hsp (human SP) and  
 CC MBP/msp (murine SP) both proteins strongly bound double stranded  
 CC RNA. Both fusion proteins also directly bound labeled ds RNAs and  
 CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the  
 CC treatment of RNA virus infections, especially those caused by  
 CC retroviruses, in particular human immunodeficiency virus.

Sequence 2859 BP; 747 A; 750 C; 712 G; 650 T; 0 other;

Query Match 37.5%; Score 1255; DB 21; Length 2859;  
 Best Local Similarity 86.7%; Pred. No. 0;  
 Matches 1419; Conservative 0; Mismatches 210; Indels 8; Gaps 3;

QY 454 aaaaacataaacccctacgtagaaactaaatgaactgcatgaaactggaanaaac 513  
 DB 264 aaaaagataagcttctaccattgtagctcaatgactgtgtggaactggaanaaac 323  
 QY 514 aatgtataagcctgtgacccctactctcgatgacgacccactaataacatgag 573  
 DB 324 aatgtataagcccgtagaccctcactcgatgacgacccactaataatgacgag 383  
 QY 574 aggaagtgcttcccccaggtactttaccattccagttccacattactatca 633  
 DB 384 tggagtgctctacccccagatactttaccattccagttccacattactatca 443  
 QY 634 agtgaacattctgtggagagacaaatltatgcaagaagaacagacagctgc 693  
 DB 444 agtgaagctctcgtggagagacaggttaatggaagaagaatgagccacccgt 503  
 QY 694 gaacacagatgctgctgcacaaagcgttgagatcctgcagaaatgagccctgcagagag 753  
 DB 504 gaacacagatgctgctgcacaaagcgttgagatcctgcagaaatgagccctgcagagag 563  
 QY 754 gctggaagtgaaatggaagaatccgaagaanaaatcccaataatcgaataatgca 813  
 DB 564 gctggaagtgaaatggaagaatccgaagaanaaatcccaataatcgaataatgca 623  
 QY 814 agtgttgagatgcatgaaacgaactgctgtaattcgaagtgagcccgagagag 873  
 DB 624 agtgttgagatgcatgaaacgaactgctgtaattcgaagtgagcccgagagag 683  
 QY 874 tggccacaccacatgaaagacttctggaacaaagttcgtgtggagagttgtgaggga 933  
 DB 684 tggccacaccacatgaaagacttctggaacaaagttcgtgtggagagttgtgaggga 743  
 QY 934 aggtgaagggaaaagcaagaagattccaagaanaatgcgcgcacagctgttcttgaga 993  
 DB 744 aggtgaagggaaaagcaagaagattccaagaanaatgcgcgcacagctgttcttgaga 803

QY 994 gctgaagaagtaccgcccctgctgcagcttgaaacagtgaaagccatgaatcaaaaaga 1053  
 DB 804 gcttagagagctgccaacccctcctgctgtgagagagtgaaagccacgaatcaagaaga 863  
 QY 1054 aaaaaaaccaatgataagcacaagaagaagcagaataatgagcagggatcaatccgat 1113  
 DB 864 aagtcagccacactgca-----agacagcccccagatatgtgcgaagatgatattctat 917  
 QY 1114 tagccagctggccagatccagcagaacaaagaagaagaagagcagagatacagctctc 1173  
 DB 918 tagtagacttgacagatccagcagaacaaagaagaagaagcagagatacagctctc 977  
 QY 1174 cacagagcagagcctcccgccgagcagagagttctgatatgcaggtgaaagttgaaacca 1233  
 DB 978 tacagaaacagagttctccacgctcgaagaggttttgatgacagtgaaaggttgagcaca 1037  
 QY 1234 cactcagaagaagaacggcgaacacaagaagaagtgccaaagcnaatgcaagcagaacat 1293  
 DB 1038 cactcagaagaagaagtggttaccaataagaagtgccaaagcnaatgcaagcagaacat 1097  
 QY 1294 gctggaagatcctgtgttcaaaagttcccgagcagagccaccacaacccgacatgaatc 1353  
 DB 1098 gctggaagatcctgtgttcaaaagttcccgagcagagccaccacaacccgacatgaatc 1157  
 QY 1354 aagaagagaagacaccacataaagaacccaagggatgaaagaagaatgaaccttttgaacc 1413  
 DB 1158 aagaagaagaagatccacataaagaacccaagggatgaaagaagaatgaaccttttgaacc 1217  
 QY 1414 tggctctggagatgaanaatgagactagtaataagaagaatgagttcagagatgcttatct 1473  
 DB 1218 tagccctgggagatgaanaatgagactagtaataagaagaatgagttcagagatgcttatct 1277  
 QY 1474 aagtcacagcagcctcccgctgcagatctccacatggtgagcagaggttcgacagagctgt 1533  
 DB 1278 tagccacacagcagcctcccgctgcagatctccacatggtgagcagaggttcgacagagctgt 1337  
 QY 1534 aggaattagtcagaagacatcacacaagaatlttaccagggcagctcgaatcctgcaca 1593  
 DB 1338 cggggttagtcaagaagacacacacaagaatlttaccagggcagctcgaatcctgcaca 1397  
 QY 1594 ggcacagtgtaactgcacatgataagccgagagttgtgtatgaggggcaactcgccacagc 1653  
 DB 1398 ggcacagtgtaactgcacatgataagccgagagttgtgtatgaggggcaactcgccacagc 1457  
 QY 1654 cgagacatltttaaagaataacatctcttcagggcagcagtgaccacaatggaactcagag 1713  
 DB 1458 cgagacatltttaaagaataacatctcttcagggcagcagtgaccacaatggaactcagag 1517  
 QY 1714 accctctgagcaactggaactatcttccagagttccagagatltccagttgaaatacaaga 1773  
 DB 1518 accctctgagcaactggaactatcttccagagttccagagatltccagttgaaatacaaga 1577  
 QY 1774 cttccccaacaaacaaagaacgaatttgtatctttatcaattgtctcctccagccac 1833  
 DB 1578 tttccccaacaaacaaagaacgaatttgtatctttatcaattgtctcctccagccac 1637  
 QY 1834 tctgatacagcaatggtatcggcaagaatgtagtgaccccgcaatgatatgagctgcgtgaa 1893  
 DB 1638 tctgatacagcaatggtatcggcaagaatgtagtgaccccgcaatgatatgagctgcgtgaa 1697  
 QY 1894 catcttaagtgtcgtgtctgagatttgaccacaagaatgacagagatgacagaagaaga 1953  
 DB 1698 catcttaagtgtcgtgtctgagatttgaccacaagaatgacagagatgacagaagaaga 1757  
 QY 1954 cggaccaatgtctgtgtgtggaaggtgtgaaactttctgagcattgacatataaa 2013  
 DB 1758 tggacaaatlttcaagctgcgggaaggtgtgaaactttctgagcattgacatataaa 1817  
 QY 2014 tcccaacataatactgaaataactgaa-aaactgtcttgaanaatttggaatttcgataacc 2072  
 DB 1818 -cccaacataatactgaaataactgaa-aaactgtcttgaanaatttggaatttcgataacc 1876  
 QY 2073 tccagtgaggccgagaga 2089

Db 1877 tccagtg9gccaagaca 1893

|||||

RESULT 7

XX X39991

XX ID X39991 standard; DNA; 1127 BP.

XX AC X39991;

XX DT 02-JUL-1999 (first entry)

DE Prostate cancer associated gene.

XX Cancer associated antigen; diagnosis; research; treatment; human;

XX breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;

XX prostate cancer; ss.

XX Homo sapiens.

XX MO904265-A2.

XX PD 28-JAN-1999.

XX PF 15-JUL-1998; 98WO-US14679.

XX PR 22-JUN-1998; 98US-0102322.

XX PR 17-JUL-1997; 97US-0896164.

XX PR 10-OCT-1997; 97US-0061599.

XX PR 10-OCT-1997; 97US-0061765.

XX PR 10-OCT-1997; 97US-0948705.

XX PR 11-OCT-1997; 97GB-0021697.

XX PA (LUDWIG) INST CANCER RES.

XX PI Chen Y, Gout I, Gure A, O'Hare M, O'bata Y, Old LJ;

XX PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;

XX PI Tureci O;

XX DR WPI: 1999-132448/11.

XX PT New isolated cancer associated nucleic acids and polypeptides -

XX PT isolated using sera from cancer patients; used to develop products

XX PT for the diagnosis, monitoring or treatment of cancers

XX PS Claim 67; Page 625-626; 787pp; English.

XX CC The invention relates to a method for diagnosing a disorder characterised

XX CC by expression of a human cancer associated antigen precursor coded for by

XX CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a

XX CC biological sample isolated from a subject with an agent that specifically

XX CC binds to the NAM, an expression product or a fragment of an expression

XX CC product complexed with an HLA molecule; and (b) determining the

XX CC interaction between the agent and the NAM or the expression product as a

XX CC determination of the disorder. The products and methods can be used in

XX CC the diagnosis, monitoring, research, or treatment of conditions

XX CC characterised by the expression of various cancer associated antigens.

XX CC The invention provides nucleic acid sequences and encoded polypeptides

XX CC which are cancer associated antigen precursors expressed in human breast

XX CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and

XX CC lung cancer.

XX SQ Sequence 1127 BP; 349 A; 265 C; 307 G; 203 T; 3 other:

Query Match 33.1%; Score 1107.2; DB 20; Length 1127;

Best local similarity 99.4%; Pred. No. 7.2e-266;

Matches 1120; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 651 gagacagcaatttaagcaagaagaacagacagctgcgaacaacagatgctgctg 710

|||||

Db 1 gagacagcaatttaagcaagaagaacagacagctgcgaacaacagatgctgctg 60

QY 711 ccaaaagcttgagatcctctcagaaatgaagccctctgccaagagaagcttgaaatgtaa 770

|||||

Db 61 ccaaaagcttgagatcctctcagaaatgaagccctctgccaagagaagcttgaaatgtaa 120

QY 771 gagatccgaagaagaataatctcaataatctgaataatgaatgaatgattggac 830

|||||

Db 121 gagatccgaagaagaataatctcaataatctgaataatgaatgaatgattggac 180

QY 831 ttaaaaggaacttgcttgaaattctgaatttgagctgcccggagagatggccaccatga 890

|||||

Db 181 ttaaaaggaacttgcttgaaattctgaatttgagctgcccggagagatggccaccatga 240

QY 891 agaacttgagcaagatctcgatgggagcttgaggggagagtgaaaggaaagaa 950

|||||

Db 241 agaacttgagcaagatctcgatgggagcttgaggggagagtgaaaggaaagaa 300

QY 951 agaagattcaaaagaataatgcgcacatagctgtcttgagagctgaagaattccgc 1010

|||||

Db 301 agaagattcaaaagaataatgcgcacatagctgtcttgagagctgaagaattccgc 360

QY 1011 ccttgctcagatgtaagcagatgaagcctgaatcaaaaagaagaacacacacatagta 1070

|||||

Db 361 ccttgctcagatgtaagcagatgaagcctgaatcaaaaagaagaacacacacatagta 420

QY 1071 agccacagacaaagcccaagatatgagcagggagatcaatccgaattagccagctggccaga 1130

|||||

Db 421 agccacagacaaagcccaagatatgagcagggagatcaatccgaattagccagctggccaga 480

QY 1131 tccagcagcgaagaaagagagagagcagatcaacgctctctcacaagcgaagcctcc 1190

|||||

Db 481 tccagcagcgaagaaagagagagagcagatcaacgctctctcacaagcgaagcctcc 540

QY 1191 cgcgcgcagggagcttgatgagcagatgaagcttgaaacacacacacacacacacacac 1250

|||||

Db 541 cgcgcgcagggagcttgatgagcagatgaagcttgaaacacacacacacacacacacacac 600

QY 1251 gcaacacagaagaagctgagcgaagcgaatgcagcgcagagaacatgctgagatccttgct 1310

|||||

Db 601 gcaacacagaagaagctgagcgaagcgaatgcagcgcagagaacatgctgagatccttgct 660

QY 1311 tcaaaagctccgcagcagcagcccaacacacacacacacacacacacacacacacacac 1370

|||||

Db 661 tcaaaagctccgcagcagcagcccaacacacacacacacacacacacacacacacacac 720

QY 1371 taaagaacacagggagcttgaaagaagaatcctttttgaactgctctgggagatgaa 1430

|||||

Db 721 taaagaacacagggagcttgaaagaagaatcctttttgaactgctctgggagatgaa 780

QY 1431 atggagctagtaataagaagagatgagttcagagatgctcttaagtcatcagacagctgc 1490

|||||

Db 781 atggagctagtaataagaagagatgagttcagagatgctcttaagtcatcagacagctgc 840

QY 1491 ctgctggaattctccatgctgcccagagctgcagcagctgtagagatgagtaagagac 1550

|||||

Db 841 ctgctggaattctccatgctgcccagagctgcagcagctgtagagatgagtaagagac 900

QY 1551 atcacccaaagatttaccagagcagctccgaatcctgcgaagcgaagcagatgagcga 1610

|||||

Db 901 atcacccaaagatttaccagagcagctccgaatcctgcgaagcgaagcagatgagcga 960

QY 1611 tgaatcccgagagctgtgtatgaggagcactgcgcccaag-cgaagacattttaag 1669

|||||

Db 961 tgaatcccgagagctgtgtatgaggagcactgcgcccaag-cgaagacattttaag 1020

QY 1670 aataaactctcacaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1729

|||||

Db 1021 aataaactctcacaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1080

QY 1730 gactatcttcagagatccagagatccagagcttcagagatccagagatccagagatccagag 1776

|||||

Db 1081 gactatcttcagagatccagagatccagagatccagagatccagagatccagagatccagag 1127

```
RESULT 8
X40002
ID X40002 standard; DNA: 946 BP.
XX
AC X40002:
XX
DT 02-JUL-1999 (first entry)
XX
DE Prostate cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9904265-A2.
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDWIG-) LUDWIG INST CANCER RES.
PI Chen Y, Gout I, Gure A, O'Hare M, Ohta Y, Old LJ;
PI Pfeundschnuh M, Sahlin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
XX WPI: 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Claim 67; Page 629-630; 787pp; English.
XX
XX The invention relates to a method for diagnosing a disorder characterised
XX by expression of a human cancer associated antigen precursor coded for by
XX a nucleic acid molecule (NAM). The method comprises: (a) contacting a
XX biological sample isolated from a subject with an agent that specifically
XX binds to the NAM, an expression product or a fragment of an expression
XX product complexed with an HLA molecule; and (b) determining the
XX interaction between the agent and the NAM or the expression product as a
XX determination of the disorder. The products and methods can be used in
XX the diagnosis, monitoring, research, or treatment of conditions
XX characterised by the expression of various cancer associated antigens.
XX The invention provides nucleic acid sequences and encoded polypeptides
XX which are cancer associated antigen precursors expressed in human breast
XX cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
XX lung cancer.
XX
SQ Sequence 946 BP; 292 A; 208 C; 259 G; 170 T; 17 other:
```

```
Query Match 23.9%; Score 798.8; DB 20; Length 946;
Best Local Similarity 93.6%; Pred. No. 3.7e-189;
Matches 888; Conservative 0; Mismatches 54; Indels 7; Gaps 6;
```

```
OY 651 gagagcagcaatttaatgcaaggaaagacagacagcgtcgcgaacaacagatgctgctg 710
    |||||||
DB 1 gagagcagcaatttaatgcaaggaaagacagacagcgtcgcgaacaacagatgctgctg 60
OY 711 ccaagcgcttagagctcgcgaatgagcccttcgcagagagcggtgagtggaatgaa 770
    |||||||
DB 61 ccaagcgcttagagctcgcgaatgagcccttcgcagagagcggtgagtggaatgaa 120
```

```
OY 771 gagaatccgaagaagaataatcctaataaactgtaaatgaatgaatgattgagattgcac 830
    |||||||
DB 121 gagaatccgaagaagaataatcctaataatctgtaaatgaatgaatgattgagattgcac 180
OY 831 ttaacggaacttgcctctgtaatttcgaggtggtcccggaagagatggtccaccaccatga 890
    |||||||
DB 181 ttaacggaacttgcctctgtaatttcgaggtggtcccggaagatggtccaccaccatga 240
OY 891 agaacttctgacccaaggttcctcggttggtggtggtggtggtggtggtggtggtggtggt 950
    |||||||
DB 241 agaacttctgacccaaggttcctcggttggtggtggtggtggtggtggtggtggtggtggt 300
OY 951 agaagattcacaagaataatccgcacatagctgttcttgaagagctgagaagattacccg 1010
    |||||||
DB 301 agaagattcacaagaataatccgcacatagctgttcttgaagagctgagaagattacccg 360
OY 1011 cccgtgctgcagttgtaacgagtaaaagcctaagaatcacaagaataaaccataagtcata 1070
    |||||||
DB 361 cccgtgctgcagttgtaacgagtaaaagcctaagaatcacaagaataaaccataagtcata 420
OY 1071 agccacagacaaagcccaagataatgcccagggatccaatccgattagccgactggtccaca 1130
    |||||||
DB 421 agccacagacaaagcccaagataatgcccagggatccaatccgattagccgactggtccaca 480
OY 1131 tccaagcagcaaaaaaagagaagaagagccagagatcacgcctcctcacaagacgagcctcc 1190
    |||||||
DB 481 tccaagcagcaaaaaaagagaagaagagccagagatcacgcctcctcacaagacgagcctcc 540
OY 1191 cgcgcgcagagaggtttgtgattgacaggtgaaagtttggaaccacac-c-gcacgaagagacg 1249
    |||||||
DB 541 cgcgcgcagagaggtttgtgattgacaggtgaaagtttggaaccacac-c-gcacgaagagacg 600
OY 1250 ggcacacaaagaagctggtccaaagcgcacatgcagccgagaaatgctgagatccttgt 1309
    |||||||
DB 601 ggcacacaaagaaggtgtggtccaaagcgcacatgcagccgagaaatgctgagatccttgt 660
OY 1310 ttc-aaagttccgcagcgcgcacccacaaaccgcactcaagtacagaagaagacac 1368
    |||||||
DB 661 ttcaaaagttccgcagcgcgcacccacaaaccgcactcaagtacagaagaagacac 720
OY 1369 cataaagaacacagaggtatgaaagaagaatgtaacctttttaacctgctctggtgagatga 1428
    |||||||
DB 721 altaagaacacagaggtatgaaagaagaatgtaacctttttaacctgctctggtgagatga 778
OY 1429 aaatgagactgaataaagaagatgagttcagaagatgcttaacttaactcaacagcagct 1488
    |||||||
DB 779 aaatgag-ctgtantaaagaagatgagttcagaagatgcttaacttaactcaacagcagc 836
OY 1489 gctgtgtggaattcttcccatggtggtccgaggtgcgcagagctgtagagttagtcagaag 1548
    |||||||
DB 837 tgtgtgtggaattcttcccatggtggtccgaggtgcgcagagctgttagagttta-ccang 895
OY 1549 acatcacacaaagatttaccagagcagctccgaaatcttcgaagagcc 1597
    |||||||
DB 896 cnttccncccaagaatttcccgagcagtttccaatctgcgaagagcc 944
```

```
RESULT 9
X40000
ID X40000 standard; DNA: 844 BP.
```

```
XX
AC X40000:
XX
DT 02-JUL-1999 (first entry)
XX
```

DE Prostate cancer associated gene.

KW Cancer associated antigen; diagnosis; research; treatment; human;  
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
KW prostate cancer; ss.

OS Homo sapiens.

```

PN W09904265-A2.
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX WPI; 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Claim 67; Page 629; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
SQ Sequence 844 BP; 251 A; 196 C; 234 G; 156 T; 7 other;

Query Match          22.6%; Score 755.6; DB 20; Length 844;
Best Local Similarity 96.3%; Pred. No. 1.9e-178;
Matches 812; Conservative 0; Mismatches 26; Indels 5; Gaps 4;

OY 813 aagtttgagattgacttaaggaacttgctgtgaatttcgagtggtgcgggaga 872
DB 1 aaggtttgagattgacttaaggaacttgctgtgaatttcgagtggtgcgggaga 60
OY 873 gtggccacccacatagaagaacttgtagcaagtttcgattggggagttgtggag 932
DB 61 gtggccacccacatagaagaacttgtagcaagtttcgattggggagttgtggag 120
OY 933 aaggtgaagggaagaagaagaatttcaagaataatgcgcatagtctgtttgaag 992
DB 121 aaggtgaagggaagaagaagaatttcaagaataatgcgcatagtctgtttgaag 180
OY 993 agctgaagaagttaacgcccctgctgctgtagtgaacgaataagcttaagcttaaaaaa 1052
DB 161 agctgaagaagttaacgcccctgctgctgtagtgaacgaataagcttaagcttaaaaaa 240
OY 1053 aaacaaacccatagtcagccacagacaagccaggaatataggccagggatcaatccga 1112
DB 241 aaacaaacccatagtcagccacagacaagccaggaatataggccagggatcaatccga 300
OY 1113 ttagccgactggccagatccagcagcagcaaaaagaagaagaagccagatcaacgctcc 1172
DB 301 ttagccgactggccagatccagcagcagcaaaaagaagaagaagccagatcaacgctcc 360

```

```

OY 1173 tcacagagcagagcctcccgccgagggagttgtgtagtcaggtgaaggttgaaac 1232
DB 361 tcacagagcagagcctcccgccgagggagttgtgtagtcaggtgaaggttgaaac 420
OY 1233 acactgcagaagaagcagcagccacaaagaagttggtccaaagcgcaatgtcagccgagaaca 1292
DB 421 acactgcagaagaagcagcagccacaaagaagttggtccaaagcgcaatgtcagccgagaaca 480
OY 1293 tctgtgagatctctgtgttcaaaagttcccgacgagccacccaaaccgcaactaaat 1352
DB 481 tctgtgagatctctgtgttcaaaagttcccgacgagccacccaaaccgcaactaaat 540
OY 1353 cagaggaagaagacacccataaagaacccagggatgtgaagaagaatgaacttttgaac 1412
DB 541 cagaggaagaagacacccataaagaacccagggatgtgaagaagaagaatgaacttttgaac 600
OY 1413 ctggtctctgggtatgaataatggagtagtaataaagaagatagttcagatgcttacc 1472
DB 601 ctggtctctgggtatgaataatggagtagtaataaagaagatagttcagatgcttacc 660
OY 1473 taagtcatcagagctgctgtctgtggaattcttcccatgtgtccggaggtcgccagagctg 1532
DB 661 taagtcatcagagctgctgtctgtggaattcttcccatgtgtccggaggtcgccagagctg 720
OY 1533 tagt-agttagtcaagagacatcacacaaagaatttaccagagcagctccgaatcctgcc 1591
DB 721 tagtcaagagacatcacacaaagaatttaccagagcagctccgaatcctgcc 779
OY 1592 aagcgaacggttaactgcataatagccagagattgtgtatgggggaacctcgccaca 1651
DB 780 angcungcng--tactgcatagtatagcc--anagtgtgtatgtgggcaantttggccca 836
OY 1652 gcc 1654
DB 837 ggc 839

RESULT 10
X39992
ID X39992 standard; DNA; 773 BP.
XX
AC X39992;
XX
DT 02-JUL-1999 (first entry)
XX
DE Prostate cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
OS Homo sapiens.
XX
PN W09904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX WPI; 1999-132448/11.

```





```
OY 2731 ttattatcagtaaacccagacaglttcagtgatgcaaatgtgtccctcgt 2790
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 791 TTTAAATTCCTGGCTAAACCCAGNACAGTCCAGATGATCAATTTGGGCCNTCTGCT 792
OY 2791 tcaagtgaac-cagtcctggaacttcaaaaacctgaataagtcctccacagltgtataa 2849
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 731 TCAGCTGAACACCGTCTCGACTTCAAAAACCTTGAATAGTCTCCCCAGTTGTATAA 672
OY 2850 attgagcaattaggaatttaacttagatgatatttgggtccattttttcatt 2909
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 671 ATTGACCAATTAAGAAATTTTAACTTTAGATGATCATTTTGTTCATTTTATNTCATTT 612
OY 2910 ttatttctgtaaatcaaacagac-ttaatgaacttgatcctgttttaagalia 2968
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 611 TTTATTTTGTATATGCAACAGAGACTTTAAATGAACCTTTGATCTCTGTTTTAAAGATTA 552
OY 2969 ttaaaaaacattgtatcatalacataatgctcttgaggacttgcttccactacac 3028
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 551 TTTAAAAACATTGTGTATCTATATGATATGGCTCTTGAGGACTTACGTTTCACTACACTAC 492
OY 3029 aggatatgctcccatgtatgctcattataaacctgcagagtgatttccagagtgcagat 3088
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 491 AGCATATGATCTCCATGATGTCATATAAACCTGACAGAGTATTTCACAGAGTCTCGAT 432
OY 3089 accttaattacatctccattagagctgaaagaatgaactagcttctgttatacagctg 3148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 431 ACCTTATATTAATCTCCATTAAGGCTGAAAGAAATGACCTTACGTTTCTTATACACTG 372
OY 3149 tgtgtcttctgagtggtgttactgtacacagaagtgtgtgacactgaggctcgtgtg 3208
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 371 TGTGTCTTTGATGTGTGTACTGTACACAGAAAGTGTGCACTGAGGCTCGCGTGTG. 312
OY 3209 gtccgatatgaaacctgttagccctgcgagtgtaagtactgcttccatttggttacg 3268
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 311 GTCCGATGAGAAAGCCTGTAGCCCTGCGAGTTAGTACTGCTTCATTCATTTGTTACG 252
OY 3269 ctggaatttctccccaatggaatgtgaatgaactaaagtgtgttgcatacaataatg 3328
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 251 CTGGAAATTTTCTCCCAATGGAATGTAAACTTAAGTGTGTCATCAATAAATGG 192
OY 3329 taactactaa 3338
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191 TAATACTAAA 182
```

```
RESULT 12
X3993/c
ID X39993 standard: DNA; 794 BP.
XX
AC X39993;
DT 02-JUL-1999 (first entry)
DE Prostate cancer associated gene.
XX
KM Cancer associated antigen; diagnosis; research; treatment; human;
KM breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KM prostate cancer; ss.
XX
OS Homo sapiens.
XX
PN MO9904265-A2.
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98MO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
```

```
PR 11-OCT-1997; 97GB-0021697.
XX
XX (LUDW-) LUDWIG INSR CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
XX MPI; 1999-132448/11.
XX
XX New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
XX Claim 67; page 626; 787pp: English.
XX
XX The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
XX Sequence 794 BP; 273 A; 141 C; 152 G; 206 T; 22 other;
```

```
Query Match 15.0%; Score 501.4; DB 20; Length 794;
Best local similarity 96.3%; Pred. No. 3,4e-115;
Matches 517; Conservative 0; Mismatches 18; Indels 2; Gaps 1;

OY 2804 tcttgagcttcaaaaacctgataagtcctcc-acagttgtataaatggaacttc 2861
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 718 TCTGGACTTTCAAAACCTTGATAAGTCTCCCNCCAGTGNATTAATGGCCNATTT 659
OY 2862 aggaatttaactttagatcatcttggttccatttattcatatttatttct 2921
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 658 AGGAATTTTAAACNTTGAATGATTCATTTGGTTCATTTTATTTATTTTGGT 599
OY 2922 aatgcaacagagacttaaatgaacttgatctctgttttaagaatlaaaaaactg 2981
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 598 AATGCCAACAAGNCTTAAATGAACCTTGATCTCGTTTAAAGATTTTAAANAACATTTG 539
OY 2982 tgtatcatacatatgctcttgaggacttagcttttaactacaactaagatatgactc 3041
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 538 TGNATCTATACATATGCGCTTGGAGACTTAGCTTACATACATACAGATTAAGATCTC 479
OY 3042 catgtagcatataaacctgcagagatgtttccagagatgcctgaactaacta 3101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 478 CATGTAGTCCATATTAACCTGCAAGATGATTTTCCANAAGTCTGATACCTTTATTTACA 419
OY 3102 tctccattagagctgaaagaatgacctagcttctgtlatacagctgtgtcttggat 3161
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 418 TCTCCNTTAGGGCTGAANAAGATGACCTACGTTCTGTATAACANCTGTGTGCTTTGAT 359
OY 3162 gtgtgttactgttaacaagaagtgtgtgacactgagagctcgtgtgtgtccgtatgaaa 3221
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 358 GTTGTGNTACTGNACACAGAGATGTGTGCMCTGAGGCTGTGCGTGTGATGTGAGAA 299
OY 3222 acctgtgagccctgcgagtgtaagtactgctccattcaattgtaagcgtgaactttct 3281
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 298 GCCTGGTAGCCCTGCGAGTTAAGTACTGCTTCCATTTATGTTTATACCTGGAATTTTCT 239
OY 3282 ccccatggaatgtaagaactaaactaaagtgttgcatacaatgtaataactaa 3338
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 CCCCATGGAATGTAAAGAACTTAAGTGTGTCATCAATTAATGTAATACTAAA 182
```

```

RESULT 13
279981/c
ID 279981 standard; cDNA; 660 BP.
XX
AC 279981;
XX
DT 07-APR-2000 (first entry)
XX
DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:65.
XX
KW Human; gene expression product; diagnosis; tumour; colon cancer;
KM colorectal adenocarcinoma; cell line SW480; cell proliferation;
KM cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
KM hyperplasia; ds.
XX
OS Homo sapiens.
XX
PN W0964576-A2.
XX
PD 16-DEC-1999.
XX
PF 09-JUN-1999; 99WO-IB01062.
XX
PR 10-JUN-1998; 98US-0088801.
XX
PA (FARB ) BAYER CORP.
XX
PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catino JU, Dertl A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;
XX
DR WPI: 2000-087220/07.
XX
PT Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer
XX
PS Claim 1; Page 158; 469pp; English.
XX
CC 279917 to Z80766 represent double stranded cDNA clones isolated from the
CC human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g. carcinomas
CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used
CC for developing agents for the diagnosis and treatment of disorders
CC involving unwanted cell proliferation, such as neoplasia, dysplasia or
CC hyperplasia.
XX
SQ Sequence 660 BP; 133 A; 172 C; 159 G; 174 T; 22 other:

Query Match 13.0%; Score 433.8; DB 21; Length 660;
Best Local Similarity 92.4%; Pred. No. 2.1e-98;
Matches 495; Conservative 0; Mismatches 36; Indels 5; Gaps 4;

```

```

DB 356 AACCGGACATCAGTCAGAGAGAGAACACCCATTAAGAAACAGGGGATGAGAAAA 297
QY 1397 gtaaccttlttgaacctgtgctgaggatgaataatgggactagtaataaaggatgag 1456
DB 296 GTAACCTTTTGAACCTGCTCTGGGGATGAATAATGGACTAGTATTAAGACATGAG 237
QY 1457 ttcaagatgccttataatcaatcagcagctgctgtgaattcttccatgggccc 1516
DB 236 TTCAGATGCTTATCTATGATCATGATCAGAGCTGCGCTGATGATTTCCATGGTGCC 177
QY 1517 gaggtgcgccagctgtgagattagtaagagacatacaaccaagaatttccagggca 1576
DB 176 GAGGTGCGCCAGCTGTAGAGATTAGTCAAGAGACATCACCAAGATTATTACAGGGCA 117
QY 1577 gctcgaatccctgcgaaggcagcaggttaactgcatgatagccggaagtgtgtatgg 1636
DB 116 GCTCCGATCTCTGCCAAGGCCACGGTAAGTGCATGATATGCCGAGAGTGTGTATGG 57
QY 1637 ggcacctgcgccacagccgagacattttaagaatacatcttcaggccagt 1692
DB 56 GGCACCTGCCCCACAGCCGAGACCAATTTAAAGATAACATCTCTTCAGCCACGT 1

RESULT 14
280269
ID Z80269 standard; cDNA; 727 BP.
XX
AC Z80269;
XX
DT 07-APR-2000 (first entry)
XX
DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:353.
XX
KW Human; gene expression product; diagnosis; tumour; colon cancer;
KM colorectal adenocarcinoma; cell line SW480; cell proliferation;
KM cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
KM hyperplasia; ds.
XX
OS Homo sapiens.
XX
PN W0964576-A2.
XX
PD 16-DEC-1999.
XX
PF 09-JUN-1999; 99WO-IB01062.
XX
PR 10-JUN-1998; 98US-0088801.
XX
PA (FARB ) BAYER CORP.
XX
PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catino JU, Dertl A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;
XX
DR WPI: 2000-087220/07.
XX
PT Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer
XX
PS Claim 15; Page 273; 469pp; English.
XX
CC 279917 to Z80766 represent double stranded cDNA clones isolated from the
CC human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g. carcinomas
CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used
CC for developing agents for the diagnosis and treatment of disorders
CC involving unwanted cell proliferation, such as neoplasia, dysplasia or
CC hyperplasia.
XX

```

SQ Sequence 727 BP; 194 A; 153 C; 191 G; 149 T; 40 other;

Query Match 12.3%; Score 412.2; DB 21; Length 727;

Best Local Similarity 95.5%; Pred. No. 5.1e-93; Mismatches 17; Indels 4; Gaps 2;

Matches 444; Conservative 0; Mismatches 17; Indels 4; Gaps 2;

QY 594 ggtacttaccattccacgttccacatttaccattatccaaagtgaacttctgtggag 653  
 Db 1 ggtacttaccattccacgttccacatttaccattatccaaagtgaacttctgtggag 60  
 QY 654 gacagcaatttaatgcaaaagaaagacaagagctgcgaaacagatgctgtcgca 713  
 Db 61 gacagcaatttaatgcaaaagaaagacaagagctgcgaaacagatgctgtcgca 120  
 QY 714 aagcgttggagctctgcgaagatgagccctgcgaaagagcgtggagtgaaagag 773  
 Db 121 aagcgttggagctctgcgaagatgagccctgcgaaagagcgtggagtgaaagag 180  
 QY 774 aatccgaagaagaataatccataatctgaataagtcagtggttggatgactta 833  
 Db 181 aatccgaagaagaataatccataatctgaataagtcagtggttggatgactta 240  
 QY 834 aacggaacttgcctgtgaatttcgaggtgcccggagagctggccacccacatgaaga 893  
 Db 241 aacggaacttgcctgtgaatttcgaggtgcccggagagctggccacccacatgaaga 300  
 QY 894 acttttggcgaaggttctggttggaggttgggtggagaaagtgaaggaagaaga 953  
 Db 301 acttttggcgaaggttctggttggaggttgggtggagaaagtgaaggaagaaga 360  
 QY 954 agattcagaagaataatg-cgcacatagctgtctgcgaagctgcgaagaaagttaccgccc 1012  
 Db 361 agattcagaagaataatg-cgcacatagctgtctgcgaagctgcgaagaaagttaccgccc 420  
 QY 1013 ctgctcgcagttgaac--gagtaaaagcctagaatcaaaaagaaa 1054  
 Db 421 ctgctcgcagttgaac--gagtaaaagcctagaatcaaaaagaaa 465  
 RESULT 15  
 280703/c  
 ID 280703 standard; cDNA; 708 BP.  
 XX  
 AC 280703;  
 XX  
 DT 07-APR-2000 (first entry)  
 XX  
 DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:787.  
 XX  
 KW Human: gene expression product; diagnosis: tumour; colon cancer;  
 KW colorectal adenocarcinoma; cell line SW480; cell proliferation;  
 KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;  
 KW hyperplasia; ds.  
 XX  
 OS Homo sapiens.  
 OS  
 PN WO9964576-A2.  
 XX  
 PD 16-DEC-1999.  
 XX  
 PF 09-JUN-1999; 99WO-1B01062.  
 XX  
 PR 10-JUN-1998; 98US-0088801.  
 XX  
 PA (FARB ) BAYER CORP.  
 XX  
 PI Endege WO, Steimann KE, Astle JH, Burgess CC, Bushnell SE;  
 PI Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JF;  
 PI Schlegel R;  
 XX  
 DR WPI; 2000-087220/07.  
 XX

PT Novel nucleic acids, used to develop products for the diagnosis and  
 PT treatment of disorders involving unwanted cell proliferation,  
 PT particularly cancers, especially colon cancer

PS Claim 15; Page 444; 469pp; English.

CC 279917 to 280766 represent double stranded cDNA clones isolated from the  
 CC human colorectal adenocarcinoma (colon cancer) cell line SW480. The  
 CC cDNA clones can be used to generate antisense oligonucleotides which  
 CC can be used for antisense therapy. Methods and products from the present  
 CC invention can be used for identifying and/or classifying cancerous cells  
 CC present in a human tumour, particularly in solid tumours, e.g. carcinomas  
 CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used  
 CC for developing agents for the diagnosis and treatment of disorders  
 CC involving unwanted cell proliferation, such as neoplasia, dysplasia or  
 CC hyperplasia.

SQ Sequence 708 BP; 226 A; 127 C; 143 G; 191 T; 21 other;

Query Match 9.8%; Score 327.8; DB 21; Length 708;

Best Local Similarity 82.8%; Pred. No. 5.1e-72; Mismatches 83; Indels 24; Gaps 12;

Matches 515; Conservative 0; Mismatches 83; Indels 24; Gaps 12;

QY 2577 ttctctttagatgatagacac--tatataaatlcatitgagaatttcc--aat 2631  
 Db 622 TTTCCTNTGATGGAATGAGCCCTGATATAAAATTTCCNTTGGAAATTTTCCCATG 563  
 QY 2632 tgcatactagtaaatgacagcttggaaact--gtcgtgacgtacttcc--aat 2686  
 Db 562 GGATCCAGAGTTAAATGAGCCAGGTGGNAACTTGTTTAAACCTTCCTTAACCAATA 503  
 QY 2687 aatctaacggac--aaagatcatatcatggtatggtgtagacattttattcgt 2744  
 Db 502 ATTCCNAACCGGCAAAAGTTCAAAACCCAGGAGGAGNGMTAAGNATTTTAATTCATG 443  
 QY 2745 actaaccc--aggacagttcagtgatgacaaatgtgtgcctct--gttcaagctgaaga- 2800  
 Db 442 NCTTACCCAGGCGCAAGTTCCAGTAAGCAAAATTTGGGGCTTCTGTGGTTACGCTGAAC 383  
 QY 2801 cagtcctgagacttcaaaaacttga-aaagtcctcca--cagttgtataaatgtgaca 2858  
 Db 382 CAGTCCGAGACTTCCAAAACCCCTGAATTAAGTCCACACAGTTGATTAATGGNCAA 323  
 QY 2859 tttagaatttaactttagatgaca-----ttgttgcattttatttaatttta 2913  
 Db 322 TTTAGGAATTTTAAACCTTTTATGATGATCCATTTGGTTCCCATTTTATTTTCAATTTTA 263  
 QY 2914 ttcttgaatgcaaacagagacttaaatgaacttgcctcgtgtttaaagattataa 2973  
 Db 262 TTTTGTGTAATGCAACAGACCTTAATGAATTTGATCTGTGTTAAAGATTATTA 203  
 QY 2974 aaacatgtgtatcatatcatatgctcgtgaggaactttagcttactactaagagat 3033  
 Db 202 AAACATTTGTATCTATACATATGATGAGCTGTGAGGAGCTTACCTTCACTACACAGAT 143  
 QY 3034 atgactccatgtagtcataataaactcgaagtgatcttccagagtgctcgatagt 3093  
 Db 142 ATGATCTCATGTAGTCAATATTAACCTGCAAGATGATTTTCCAGATGCTCGATCTGT 83  
 QY 3094 taattacatcctatgagctgaaaagaaatgacactcgttctgtataag--ctgctc 3152  
 Db 82 TAATTACATCTCCATTGAGGCTGAAAAAATGACCCAGCTTCTGTATACAGCTGTGT 23  
 QY 3153 gctttgagtgctgtactgt 3174  
 Db 22 GCTTTGATCTGTGTACTGT 1

Search completed: April 5, 2001, 08:32:35  
 Job time: 59399 sec

---



Best Local Similarity 12.5%; Pred. No. 4,1e-10;  
Matches 41; Conservative 171; Mismatches 116; Indels 0; Gaps 0;

Oy	2	cctccbcgcggagcgcgagcgcttagcgacttcgaagtcttgacgcgagcgtcgagc	61
Db	1193	yy	1252
Oy	62	tctctgacctccgcgtctccctttgaccgctcccccccgcgcgcgcgcgcctcc	121
Db	1253	yyy	1312
Oy	122	tccacggcacctccgcctcttccctcccttcgctcccttccctccctttcttc	181
Db	1313	yy	1372
Oy	182	tctcttccctccctcgcgcgcacgccgcaggacgcgcgcgcgcgcgcgcgcgcga	241
Db	1373	yy	1432
Oy	242	gcagccagagtltaataccaactaaactctcagaactgaacaaagacaatyltcc	301
Db	1433	yxtgaccaaatcttctaattcttttacactgcatgatgatgtaattacagtgatgcc	1492
Oy	302	ggaagccctctttttaaaaaagtga	329
Db	1493	ttacatgocgitttttganaactganatga	1520

RESULT 2  
 US-08-232-463-14/c  
 Sequence 14, Application US/08232463  
 Patent No 5670367  
 GENERAL INFORMATION:  
 APPLICANT: DORNER, F.  
 APPLICANT: SCHEIFLINER, F.  
 APPLICANT: FALKNER, F. G.  
 TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
 NUMBER OF SEQUENCES: 52  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Foley & Lardner  
 STREET: 1800 Diagonal Road, Suite 500  
 CITY: Alexandria  
 STATE: VA  
 COUNTRY: USA  
 ZIP: 22313-0299  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/232,463  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/07/935,313  
 FILING DATE:  
 APPLICATION NUMBER: EP 91 114 300.6  
 FILING DATE: 26-AUG-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BENT, Stephen A.  
 REGISTRATION NUMBER: 29,768  
 REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703)836-9300  
 TELEFAX: (703)683-4109  
 TELEX: 899149  
 INFORMATION FOR SEQ ID NO: 14:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 7218 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear

;  
; IMMEDIATE SOURCE:  
; CLONE: PTZgpt-Fls  
US-08-232-463-14

Query Match 1.9%; Score 68; DB 1; Length 7218;  
Best Local Similarity 4.3%; Pred. No. 1.9e-08;  
Matches 17; Conservative 231; Mismatches 146; Indels 0; Gaps 0;

[illegible]

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525











```

APPLICANT: Fu, Ying-Hui
APPLICANT: Yu, Chang-Eh
APPLICANT: Oshima, Junko
APPLICANT: Mulligan, John T.
APPLICANT: Schellberg, Gerald D.
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
TITLE OF INVENTION: WERNER'S SYNDROME
NUMBER OF SEQUENCES: 209
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US//08/781,891
FILING DATE: 27-DEC-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: NO. 6090620tenburg P.H.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 240052.419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 87350 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-781-891-79

Query Match          1.2%; Score 41; DB 3; Length 87350;
Best Local Similarity 56.2%; Pred. No. 2.5;
Matches 77; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

OY      62 tctctcgatccgcgacttccttgtagcgcgtccccccggcgagcgagccgcgtccc 121
        ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 52222 TCCTTCCTCCCTTCCCTTCCTTCCTTCCTTCCTTCCTGCGCGAGACGCCCTGCCTTCC 55281

OY      122 tcgaacgcgaactcgcgcctcttccctcccttcgctcccttccttccttccttc 181
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 55282 TTCCTTCCTTCCTCCCTCCCTCCCTCCCTCCCTCCCTTCCTTTCTTCCTTCCTTCTTCT 55341

OY      182 ttctctccctccctccgc 198
        ||| ||| | | | | | | | |
Db 55342 TTCCTTCCTCTCTCTCTC 55358

RESULT 14
US-08-149-096A-1/C
Sequence 1, Application US/08149096A
Patent No. 5556956
GENERAL INFORMATION:
APPLICANT: ROY, A.K.
APPLICANT: CHATTERJEE, B.
TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO THE
TITLE OF INVENTION: ANDROGEN RECEPTOR GENE AND USES THEREOF
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
City: Houston
STATE: Texas
COUNTRY: USA

```

```

ZIP: 777210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/149,096A
FILING DATE: No. 555695member 4, 1993
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Denise L. Mayfield
REGISTRATION NUMBER: 33,732
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 320-7200
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2775 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: oligonucleotide
US-08-149-096A-1

Query Match      1.2%; Score 40.8; DB 1; Length 2775;
Best Local Similarity 47.6%; Pred. No. 0.33;
Matches 120; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 108 cggcgccgcctcctccagcagccactccgaccttcttcctccttgctccctttcctct 167
    ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2606 CGGCCCCCGCGGGCCCTCCCTTCCTTCCACGCCCTTTCTCTCTCCTCCCTCCCTGC 2547
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 168 cccttttcccttccttcctcccctccgcgcacccgcagaccgcgcgcgcgcgcgcgc 227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2546 CCCTCCCTCCCTCCCTTCTCTTCTTCTCCCTCCCTCACCGCAAGTAGGCGAGGTGGCTT 2487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 228 acgaatcggaagcaagaagatttaaaccaacttaaccttcagaaactgaaacaaag 287
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2486 TAGAAACAATAGTCTGGCGGTATGATTGAAGCTGGAAGAAGACTCCCTTGAGCTCTG 2427
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 288 aaacaattgttcygaagccctctltttaaaaaagtlagaacttagaactcataga 347
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2426 AAGAGCCCTTTCGCCCAATAACCACTGTCTTTCGGAACCGGGCTTTGAGACTCCAAGCA 2367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 348 ctgaattaaact 359
    ||| ||| |||
DB 2366 GAAGCTTAACCT 2355

RESULT 15
US-08-147-777-3
Sequence 3, Application US/08147777
Patent No. 5914265
GENERAL INFORMATION:
APPLICANT: Roop, Dennis R.
APPLICANT: Rothmager, Joseph A.
APPLICANT: Greenhalgh, David A.
APPLICANT: Yuspa, Stuart H.
TITLE OF INVENTION: KERATIN K1 EXPRESSION VECTORS
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: LYON & LYON
STREET: 611 West Sixth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90017
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2001, 00:29:42 : Search time 11371 Seconds  
(without alignments)  
2160.608 Million cell updates/sec

Title: US-09-316-048-5  
Perfect score: 3506  
Sequence: 1 actccgcgcggcgtcgcg9.....taatactaaaaaaaaa 3506

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues  
Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: em\_est4a:\*  
42: em\_estfun:\*  
43: em\_esthum1:\*

44: em\_esthum2:\*  
45: em\_esthum3:\*  
46: em\_esthum4:\*  
47: em\_esthum5:\*  
48: em\_esthum6:\*  
49: em\_esthum7:\*  
50: em\_esthum8:\*  
51: em\_esthum9:\*  
52: em\_esthum10:\*  
53: em\_esthum11:\*  
54: em\_esthum12:\*  
55: em\_esthum13:\*  
56: em\_esthum14:\*  
57: em\_esthum15:\*  
58: em\_esthum16:\*  
59: em\_esthum17:\*  
60: em\_esthum18:\*  
61: em\_esthum19:\*  
62: em\_esthum20:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estov1:\*  
68: em\_estov2:\*  
69: em\_estpl1:\*  
70: em\_estpl2:\*  
71: em\_estpl3:\*  
72: em\_estpl4:\*  
73: em\_estpl5:\*  
74: em\_estro1:\*  
75: em\_estro2:\*  
76: em\_estro3:\*  
77: em\_estro4:\*  
78: em\_estro5:\*  
79: em\_estro6:\*  
80: em\_estro7:\*  
81: em\_estro8:\*  
82: em\_estro9:\*  
83: em\_estro10:\*  
84: em\_estro11:\*  
85: em\_estro12:\*  
86: em\_estro13:\*  
87: gb\_est41:\*  
88: gb\_est42:\*  
89: gb\_est43:\*  
90: gb\_est44:\*  
91: gb\_est45:\*  
92: gb\_est46:\*  
93: gb\_est47:\*  
94: gb\_est48:\*  
95: gb\_est49:\*  
96: gb\_est50:\*  
97: gb\_est51:\*  
98: gb\_est52:\*  
99: gb\_est53:\*  
100: gb\_est54:\*  
101: gb\_est55:\*  
102: gb\_est56:\*  
103: gb\_est57:\*  
104: gb\_est67:\*  
105: gb\_est68:\*  
106: gb\_est69:\*  
107: gb\_est70:\*  
108: gb\_est71:\*  
109: gb\_est72:\*  
110: gb\_est73:\*  
111: gb\_est74:\*  
112: em\_esthum21:\*  
113: em\_esthum22:\*  
114: em\_esthum23:\*  
115: em\_estom1:\*  
116: em\_estom2:\*

117: em\_estp16:\*  
 118: em\_estp17:\*  
 119: em\_estp18:\*  
 120: em\_estp19:\*  
 121: em\_estp20:\*  
 122: em\_estp21:\*  
 123: em\_estp22:\*  
 124: em\_estp23:\*  
 125: em\_estp24:\*  
 126: em\_estp25:\*  
 127: em\_estp26:\*  
 128: em\_estp27:\*  
 129: em\_estp28:\*  
 130: em\_estp29:\*  
 131: em\_estp30:\*  
 132: em\_estp31:\*  
 133: em\_estp32:\*  
 134: em\_estp33:\*  
 135: em\_estp34:\*  
 136: em\_estp35:\*  
 137: em\_estp36:\*  
 138: em\_estp37:\*  
 139: em\_estp38:\*  
 140: em\_estp39:\*  
 141: em\_estp40:\*  
 142: em\_estp41:\*  
 143: em\_estp42:\*  
 144: em\_estp43:\*  
 145: em\_estp44:\*  
 146: em\_estp45:\*  
 147: em\_estp46:\*  
 148: em\_estp47:\*  
 149: em\_estp48:\*  
 150: em\_estp49:\*  
 151: em\_estp50:\*  
 152: em\_estp51:\*  
 153: em\_estp52:\*  
 154: em\_estp53:\*  
 155: em\_estp54:\*  
 156: em\_estp55:\*  
 157: em\_estp56:\*  
 158: em\_estp57:\*  
 159: em\_estp58:\*  
 160: em\_estp59:\*  
 161: em\_estp60:\*  
 162: em\_estp61:\*  
 163: em\_estp62:\*  
 164: em\_estp63:\*  
 165: em\_estp64:\*  
 166: em\_estp65:\*  
 167: em\_estp66:\*  
 168: em\_estp67:\*  
 169: em\_estp68:\*  
 170: em\_estp69:\*  
 171: em\_estp70:\*  
 172: em\_estp71:\*  
 173: em\_estp72:\*  
 174: em\_estp73:\*  
 175: em\_estp74:\*  
 176: em\_estp75:\*  
 177: em\_estp76:\*  
 178: em\_estp77:\*  
 179: em\_estp78:\*  
 180: em\_estp79:\*  
 181: em\_estp80:\*  
 182: em\_estp81:\*  
 183: em\_estp82:\*  
 184: em\_estp83:\*  
 185: em\_estp84:\*  
 186: em\_estp85:\*  
 187: em\_estp86:\*  
 188: em\_estp87:\*  
 189: em\_estp88:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	1757.8	50.1	1801	146	U69197 U69197 Soar
2	701	20.0	1018	136	BE867390 BE867390
3	646.2	18.4	1036	137	BE869880 BE869880
4	621.2	17.7	724	97	AM952360 EST364430
5	617.2	17.6	752	137	BE894295 BE894295
6	606.4	17.3	821	107	BE379687 BE379687
7	604	17.2	609	107	BE380123 BE380123
8	603.8	17.2	686	110	BE613972 BE613972
9	602.2	17.2	674	38	AV707406 AV707406
10	594.4	17.0	616	111	BE674157 BE674157
11	588.2	16.8	594	40	AM162002 AM162002
12	571.4	16.3	612	38	AV695897 AV695897
13	565.4	16.1	749	110	BE617223 BE617223
14	555	15.8	609	136	BE810330 BE810330
15	553.6	15.7	615	134	BE082712 RC2-BT064
16	550	15.7	550	107	BE348298 BE348298
17	548	15.6	564	40	AM163206 AM163206
18	546.4	15.6	829	106	BE300155 BE300155
19	542.4	15.5	584	136	BE813737 RC1-BM003
20	534	15.2	673	135	BE781223 BE781223
21	533.2	15.2	536	136	BE837505 BE837505
22	531	15.1	866	136	BE868328 BE868328
23	530.2	15.1	573	3	AA206573 AA206573
24	517.8	14.8	910	138	BE966164 BE966164
25	513	14.6	543	134	BE089861 RC5-BT070
26	511.4	14.6	512	27	AI983007 w146902.x
27	510.4	14.6	512	27	AI983007 w146902.x
28	508.8	14.5	852	137	BE872137 BE872137
29	500.6	14.3	673	109	BE541462 BE541462
30	497.4	14.2	633	89	AM320695 AM320695
31	488.2	13.9	569	3	AA191622 AA191622
32	487	13.9	516	25	AI819766 w142a10.x
33	483.4	13.8	485	25	AI802592 AI802592
34	478.4	13.6	480	134	BE042598 BE042598
35	475	13.5	494	89	AM351909 AM351909
36	470.4	13.4	509	142	N31181 N31181
37	470	13.4	517	7	AA427366 AA427366
38	470	13.4	517	92	AM580332 AM580332
39	467.4	13.3	703	137	BE888420 BE888420
40	457	13.0	724	38	AV700494 AV700494
41	455.2	13.0	561	141	H16250 H16250
42	451.8	12.9	455	96	AM874257 h96904.x
43	450.4	12.8	452	17	AI190464 q38b02.x
44	446.4	12.7	586	141	H11397 H11397
45	445.8	12.7	524	96	AM877286 MR4-PT005

## ALIGNMENTS

RESULT 1  
 LOCUS U69197 1801 bp mRNA  
 DEFINITION U69197 Soares Infant brain 1NTB Homo sapiens CDNA clone 22368, mRNA  
 ACCESSION U69197  
 VERSION U69197.1 GI:2739420  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 1801)  
AUTHORS Volorio,S., Simon,G., Repetto,M., Cucciard,M., Banfi,S., Borsani  
,G., Balabio,A. and Zollo,M.  
TITLE Sequencing analysis of forty-eight human image cDNA clones similar  
to drosophila mutant protein  
JOURNAL DNA Seq. 9 (5-6), 307-315 (1998)  
MEDLINE 99452388  
COMMENT Contact: Zollo, Massimo  
Telethon Institute of Genetics and Medicine  
Via Olgettina 58, Milan, MI 20132, Italy  
Email: zollo@tigem.it.

FEATURES  
source  
1.1801  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="19p13.1"  
/clone="22368"  
/clone.lib="Soares infant brain INTB"  
/sex="female"  
/dev\_stage="73 days post natal"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: whole brain; Vector: lafmid BA; Site.1: Not  
I; Site.2: Hind III; 1st strand cDNA was primed with a Not  
I - oligo(dT) primer (5',  
AAGTGAAGAATTCGGCGCGCAGCAATTTTATTTT 3';  
double-stranded cDNA was ligated to Hind III adaptors  
(Pharmacia), digested with Not I and directionally cloned  
into the Not I and Hind III sites of the lafmid BA vector.  
Library went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 503 a 409 c 386 g 503 t  
ORIGIN

Query Match 50.1%; Score 1757.8; DB 146; Length 1801;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1770; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1735 gctcgaatctcgcgaagccagcagcgaatcgtacatagcccgagagtgctgtatgag 1794  
Db 5 GCTCCGAATCTCTCCAAAGGCGACGTAACGTGCATGATACCCGAGAGTGTGTATGGG 64

QY 1795 ggcacctgcgcacagcccgagacattttaagaataacatctcttcaggccagtaacc 1854  
Db 65 GGCACCTCGCCACACCGCGAGCATTTTAAAGAAATAACATCTCTCAAGCCACGTAACC 124

QY 1855 catgagaccttcagagacacctctgagcaactggaactatcttcagagtcagaggatc 1914  
Db 125 CATGGAACCTCTCAGAGACCCCTCTGAGCAACTGGAATATTCTTCCAGAGTCAGGGATTTC 184

QY 1915 cagggtgaatataaagactcccaaaaacaagaagaacgaattgtatctctatcaat 1974  
Db 185 CAGGTGAATATCAAAAGCTTCCCAAAAACAACAAGAACGAATTTGTATCTTATCAAT 244

QY 1975 tgcctcttcagcacctctgatacgaatggtatcgcgcaagagatgtggagtcctgcat 2034  
Db 245 TGCCTCTTCAGCACCTCTGATCAGCCATGATCGCGCAAGAGATGTGGAGTCTCTGCAT 304

QY 2035 gatatgctgcgtgacacatcttaagttgctgtctgagttgagccacaagtatcacag 2094  
Db 305 GAATATGGCTGCGCTGAACATCTTAAAGTTGCTGTCTGAGTTGGACCAACAAGATCAGAG 364

QY 2095 atgccaagaacaggaacaggaacatgctcgtgtgtgtgagagtgctgtaacctttctgc 2154  
Db 365 ATGCCAAGAACAGGAACGGACAATGTGTGTGTGGAGGTGTGAACCTTTTCTGGGC 424

QY 2155 catgaacatataaatacccaatatatactgaaatactgaaactgcttggaaatc 2214  
Db 425 CATGAACCATATATAAATCCCAACATATATACGAATAATCTGAAACTGCTTGAATAATT 484

QY 2215 tggaaattctgataacctccagctggccgagagacaagctgtgtgaagagctgtggcaga 2274  
Db 485 TGGAAATTTTGTGATACCTCCAGATGGGCGGAGACACAGCTGGGTAAAGATGTGGCGACGA 544

QY 2275 gcaaggagaacacaagaacaagaagagcgctgtggcc-gctgtgactgtctgtgggt 2333  
Db 545 GCAGGGGAAGACAACAGAAACACAAGGAGCGGCTGTGGCGGGGTGACATGTGTGGGGT 604

QY 2334 ttgtgtgtagcgcacacgcggtgagcctgtgggttcctcctacgcaataagagcgctgtgg 2393  
Db 605 TTGTGTGATGGCCACTCGGTGACTGGCGGTCCCTTACGCAATGACGTGCTGTGGGG 664

QY 2394 aagaagcgctgcgcacagcgaactgtgtctcccggaacacacagatccacacctgtggca 2453  
Db 665 AAGAAAGCTGCGCCAGCAGCAGCTGTTCTCCCGGACACAGATCCACACCTGGGCA 724

QY 2454 cctccgtgtgtgtccttttttccctcgtgtgaagaagaacgcgacagaccctctc 2513  
Db 725 CCTCCGCTTGGCTCTTTTTCCTTCCCTGTGTGAAGAAAGACGGCAGACCCCTTCTC 784

QY 2514 aagctgctcactcagacacatltgggacaacctgtgacagacatgccaagagagagcct 2573  
Db 785 AAGCTGAGCTCACTGACACATTTGGACAAACCTTGAGACCCATGSCAGAGAGAGGCT 844

QY 2574 ttgacggcccccagagcgtlaaagcaccagagaataatcaatgtcttcctactcagctgac 2633  
Db 845 TTGACGGCGCCCGAGAGGCTAAAGACACAGAAATCAATTCCTTCTACTCACCGTGAC 904

QY 2634 ccaactttctagtgtgccaagcgcacacacacacacacacacacacacacacacacacac 2693  
Db 905 CCAACTTTTCTAGTGTCCACAGCGCCACACACCTCCGTCGATGCCACACCATCACACT 964

QY 2694 gcttcctctccacaagctgtgtgtctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2753  
Db 965 GCTTCTCTTCCACAGT 1024

QY 2754 acaatataaaatttcatlttgagaaattctcaatgtatctagttaaatagacagtt 2813  
Db 1025 ACACTATATATAAATTTTCATTTTGGAATTTCTCAATTTGTATGTAATATAGCAGAGTT 1084

QY 2814 tggaaactgtcctgagacgtacttcaatcaatcaatcaatcaatcaatcaatcaatcaat 2873  
Db 1085 TGGAAACTTGTCTGAGACTCACTTATCAATTAATCTTAACGACAAAGATATATCCATGT 1144

QY 2874 gtaatgtgttagacatttatttatttcaatgaacttcaacgaagaagttcaatgtatgaat 2933  
Db 1145 GTATGTGTTAGACATTTTATTTTCAATTGACTTAACCGAGACAGTTTCAAGTATCAAT 1204

QY 2934 tgtgtgcctctgtgtcagctgaaacagctcgtgacttccaacacacacacacacacacac 2993  
Db 1205 TGTGTGCCCTGTGTTCAAGCTGAAGAGTCTCTGACCTTTCAAAAACCTTGATTAAGTCTC 1264

QY 2994 ccaagttgtataaatgtgaacatttgaacatttgaacatttgaacatttgaacatttgaac 3053  
Db 1265 CCACAGTGTATATAATTTGCAATTTAGCAATTTTAACTTTAATATATATTTTGGTTCC 1324

QY 3054 attttatttcaatttatttatttatttgaatgaacgaagaacttgaatgaacttgaact 3113  
Db 1325 ATTTTATTTTCAATTTTATTTTGTGTATGTGAACAGAGCTTAATATATATTTGATCTCT 1384

QY 3114 gttttaaagatlaataaaacacattgtglatcctatacatalgtgctcttgaggactgact 3173  
Db 1385 GTTTTAAAGATTTATTAATAAACAATTTGTATCTATATATATATATATATATATATATAT 1444

QY 3174 ttaactaacacaggaatgatacgcagtgtagtgcataataacacacacacacacacacac 3233  
Db 1445 TTCACTACACTACAGAT 1504

QY 3234 cagagtgctcgtatgactgttataacacacacacacacacacacacacacacacacacac 3293  
Db 1505 CAGAGTCTCTGATATCTGTAT 1564

QY 3294 ctgtatacagctgtgtccttctgtatgtgtgtacgtacacagaagtggtgtcactga 3353

Db	1565	CHGTATACAGCTGTGTGCTTTTGTGATGTTGTGTACTGTACACAGAAAGTGTGGCACTGA	1624
Qy	3354	ggctctgcgtgtgtctcgtatcgtgaataaaccttgcctgcctgcagttgaactgtcttca	3413
Db	1625	GGCTGTGCGTGTGTCCTCCATGTGAAAGCGCTGGTAGCCCTGCAGGTAACTACTGCTTCCA	1684
Qy	3414	ttcatgttttaagcctcggaaattttctccccaatggaaatgaaactaaactgaagttgt	3473
Db	1685	TTCAATGTTTACGCTGGAAATTTTTCCTCCCAATGGAATGTAAAGCTTAAGTGTGTGT	1744
Qy	3474	catcaataatgtaataactaaaaaataaataa	3506
Db	1745	CATCAATAATGTATATCTATAAAAAAAAAAAAA	1777
RESULT	2		
LOCUS	BE667390	1018 bp	mRNA EST 27-SEP-2000
DEFINITION	601442470701 NIH_MGC_65 Homo sapiens CDNA clone IMAGE:3846689 5',		
ACCESSION	BE667390		
VERSION	BE667390.1	GI:10316166	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 1018)		
JOURNAL	NIH-MGC <a href="http://www.ncbi.nlm.nih.gov/MGC/">http://www.ncbi.nlm.nih.gov/MGC/</a> .		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Tel: (301) 496-1550		
	Email: Robert.Strausberg@nih.gov		
	Tissue Procurement: ATCC		
	CDNA Library Preparation: Life Technologies, Inc.		
	DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	Cloning by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:		
	<a href="http://image.llnl.gov">http://image.llnl.gov</a>		
	Plate: L10C546 row: 1 column: 18		
	High quality sequence stop: 654.		
FEATURES	Location/Qualifiers		
source	1..1018		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone_image="3846689"		
	/clone_1lb="NIH_MGC_65"		
	/tissue_type="adenocarcinoma"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: colon; Vector: pcMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.8 kb. Library constructed by Life Technologies."		
BASE COUNT	295 a 267 c 262 g 194 t		
ORIGIN			
Query Match	20.0%; Score 701; DB 136; Length 1018;		
Best Local Similarity	90.0%; Pred. No. 2.6e-163;		
Matches 809; Conservative	0; Mismatches 80; Indels 10; Gaps 5		
Qy	1280	tggtcccaatcagcagcgcaaaaagggagagcagcagatcacagctcctccacagagc	1339
Db	1	TGGCCCAATCTCAGAGCGCAAAAAGGAAGAGAGCAAGTAAACAGCTCTCACAGAGC	60
Qy	1340	gagagctcccgcgcgcgaggggttctgtatcaggtgaaggttggnaacacacacgtcgag	1399
Db	61	GAGGCTCTCCGCCGCCGACGAGGAGTTGTGTATCAGGTGAAGGTTGGAACACACACTGCGAG	120
Qy	1400	aaggaacgggacacacaagaaggtgtgcacaagcgcaatgcagctcgagaacatgtctgaga	1459

D	b		121	AAGGACGGGCACCAACAAGAAGTGGCCAAAGGCCATTCGACGCCGAGAACATCTGTGAGA	180
O	y		1460	tcccttggttccaagaagtcgccgacgcggcaccacaaccgcgaactcaagtcagaagaga	1519
D	b		181	TGCTTGTTCAAAGTCCCCGAGCGCAGCCACCACCAACCGCACACTCAAGTCAGAGAGA	240
O	y		1520	agacacccataagaanaaccaggagtatgaaagaagaagtaacctttttgaacctcgccctg	1579
O	y		1580	ggagatgaatacyggagactgaataaagaaga tgaattcagaatgaccttatctaagtcac	1639
D	b		301	GGGATGAATAATGGGACTGTATTAAGAGAGATGAGTTCCAGATGCCATTATCTAAGTCATC	360
O	y		1640	agcagctgcctcctgtgaattcctcccactgtygccgaagtcgcaccaagcctgtagaatta	1699
D	b		361	AGCAGCTGCTGCTGGAAATTCCTCCATGAGTGGCCGAGGTGCCCAAGCTGTAGAGATT A	420
O	y		1700	gtcaaagacatcacaccaagaattttaccagaagcagctccgaatcctgcgaagyccacgg	1759
D	b		421	GTCAAGGACATCACACCAAAATTTTACAGAGGACGCTCCGAATCTGCGCAAGGCCACGG	480
O	y		1760	taactgcgatgatagccccgaagatcttgtatagtgaggccactctgcaccaagccagacca	1819
D	b		481	TNACTGCCATATAGCCCAGAGATTGTGTAGTGGGGCACTCGCCACAGCCGAGCCA	540
O	y		1820	ttttaagaataacatctctctcagagccagtcgaccactgaacctctcagaagaccctctg	1879
D	b		541	TTTTNAAGATTAACATCTCTTCACAGCCACGACCCCATTTGACCTCTACAGAACCTCTG	600
O	y		1880	agcaactggaactatcttccagaagtcagaagatlcasagltgaatlacaaaagactlccca	1939
D	b		601	AGCAACTGAGCTATCTTTCACAGATCCAGAGATTCACAGTTTAATCAAGACTTCCCA	660
O	y		1940	aaacaacaagaagaatttgtact---cttatcaattgtctctccagccaccctctga	1996
D	b		661	AAAATCACCAAGAACGAAATTCGATCTCTTATCAATAGCTCTCTTCACCCACTGTAT	720
O	y		1997	tcagcatcgtatcagcgaagaagtgtggatctctccca---tgatatgctcgcgtacga	2053
D	b		721	CCACATCTGTTTTCGGCAGGTTTGCGAATCTCGGCAATGATCATGAGTGGGCTGGACA	780
O	y		2054	tcttaaag-ctgctcgtctgaatttgagacaacaaa-gtacagagatgccaagaacaggaaa	2111
D	b		781	TCTTTAAGATGGCTGGCGGAGTAGTGACCCACAAAGGTCMAAAATTTCCAGAAAGGGAA	840
O	y		2112	cggaccaaagtcgtgtgtgtgg--aggtgtcgaacctttcttcgcatcgaacattata	2168
D	b		841	GGGACCCGAGTTCCGGTGTGTTGGCAGAGGTCCTGAAACCAATTTTGGGCTCGAACCATTA	899

RESULT 3

BE889880	1036 bp	mRNA	EST	29-SEP-2000
LOCUS	60151213.f01	NIH_MGC_71	Homo sapiens cDNA clone IMAGE:3913590	5'
DEFINITION	mRNA sequence.			
ACCESSION	BE889880			
VERSION	BE889880.1	GI:10347645		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
AUTHORS	M (bases 1 to 1036)			
JOURNAL	NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.			
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Contact: Robert Strausberg, Ph.D. Tel.: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc.			



cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/ILNL at:  
<http://image.llnl.gov>

Plate: LINC720 Row: p Column: 07  
 High quality sequence stop: 706.

## FEATURES

source

Location/Qualifiers  
 1..1036  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="3913590"  
 /clone\_lib="NH\_MGC\_71"  
 /tissue\_type="Telomysarcoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: uterus; Vector: pCMV-Sport6; Site 1: NotI;  
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 2.1 kb."  
 BASE COUNT 296 a 230 c 220 g 290 t  
 ORIGIN

Query Match 18.4%; Score 646.2; DB 137; Length 1036;  
 Best Local Similarity 91.5%; Pred. No. 1.1e-149;  
 Matches 740; Conservative 0; Mismatches 63; Indels 6; Gaps 5;

QY 2579 cggccccaagctaaagaaccagagaagaatcaatgcttcctcactcagcgtgacccaac 2638  
 |||||  
 DB 1 CGGCCCCAGGCTAAAGACGACGAGAAATCAATGCTTCCTACGACGTCGACCCAAC 60  
 |||||  
 QY 2639 ttcttagtctgccaagcgcacacacccctcctcagctacccacacacacacacacac 2698  
 |||||  
 DB 61 TTTTCTAGTGTCGACGCGCCGACGCTCTGACATACCCACATCAGCCACTGCTTT 120  
 |||||  
 QY 2699 cctctccacaagctgctgcttcttcttcttcttcttcttcttcttcttcttcttctt 2758  
 |||||  
 DB 121 CCGTTCACACAGATGATCTTATTTCTTAGTTTCAATTTCTTTGATGATGACACT 180  
 |||||  
 QY 2759 atataaaatttcattcttggaattctcactatgactatgactaataagacagttcgaa 2818  
 |||||  
 DB 181 ATATATAAATTTTCATTGGAATTTCTCAATTTCTATCTAGTTAATAGCAGATTGGAA 240  
 |||||  
 QY 2819 actgtctgagaagccttctatcaataatcaaccgacaaagatcatalcagctgctg 2878  
 |||||  
 DB 241 ACTTGTCTGAGACTGACTTATCAATTAATCAACGACAAAGATCATATCATGTGTATG 300  
 |||||  
 QY 2879 tggtagaacttttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2938  
 |||||  
 DB 301 TGGTTAGACATTTTATTTTCAATTTGACTTAACCCAGACAGTTTCAAGATGCAAAATGTGT 360  
 |||||  
 QY 2939 gccctctgtctcagctggaacagctcctggaacttcaaaaaccttgataagctccaca 2998  
 |||||  
 DB 361 GCCCTCTGTTGACGTCGAACAGCTCTGACACTTTCAAAAACCTTGATAGTGTCCACA 420  
 |||||  
 QY 2999 gtctgataaattggaacattggaattttaaactttagatgatacttctgctcgtttt 3058  
 |||||  
 DB 421 GTTGTTAATATGACAAATTTAGGAATTTAAACCTTTAGATGATCAATTTGCCATTTT 480  
 |||||  
 QY 3059 tatttcaatttatttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 3118  
 |||||  
 DB 481 TATTTTCATTTTATTTTATTTTATGCAACAGACTTAATGAACTTGTATCTCTCTTTT 540  
 |||||  
 QY 3119 aaagattataaaaacattgtgtatctatacatatgagctctggaagacttaagcttcc 3178  
 |||||  
 DB 541 AAAGATTATTAATAAACATTTGTTATCTATACATATGCGCTCTTGAGAGACTTACCTTTCA 599  
 |||||  
 QY 3179 tacactacagatgatctcc-atgtaglccatataaaccctgcaagatg--atttcca 3235  
 |||||  
 DB 600 TACACTACAGGATATGATCTCCAAATGATGATCAATTAACCTGCAAGATGATGATTTCCAG 659  
 |||||  
 QY 3236 gagtctcagatct-gtataatacctccatataagctggaagaagataagctcagcttcc 3294  
 |||||  
 DB 660 AAGTGTGATACATGCTGATTTACTCTCATTTAGGATGAAGAAATGACACTAGTTCTG 719  
 |||||

QY 3295 tctataacgctgtgtcttcttcttcttcttcttcttcttcttcttcttcttcttctt 3353  
 |||||  
 DB 720 TATAAAGTGTGTTGCTTTCAGGAGTTGGCGGTACTGAGAACGAGAGTGTTCCTCTGA 779  
 |||||  
 QY 3354 ggcctcgt 3382  
 |||||  
 DB 780 GGGCTGTGATGTGGCGCGCTGTGTATACC 808  
 |||||

## RESULT 4

AM952360 724 bp mRNA EST 01-JUN-2000  
 LOCUS EST564430 MAGE resequences, MAGB Homo sapiens cDNA, mRNA sequence.  
 DEFINITION AM952360  
 ACCESSION AM952360  
 VERSION AM952360.1 GI:8142042  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE  
 1. (bases 1 to 724)  
 Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt, I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and Quackenbush, J.

Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray  
 Unpublished (2000)  
 CONTACT: John Quackenbush

The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 3528  
 Fax: 301 838 0208  
 Email: johnd@tigr.org

Plate: 38  
 Seq primer: Reverse.

## FEATURES

source

Location/Qualifiers  
 1..724  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="MAGE resequences, MAGB"  
 /note="Vector: pBluescriptSkm"  
 BASE COUNT 198 a 179 c 189 g 158 t  
 ORIGIN

Query Match 17.7%; Score 621.2; DB 97; Length 724;  
 Best Local Similarity 95.1%; Pred. No. 1.7e-143;  
 Matches 685; Conservative 0; Mismatches 28; Indels 7; Gaps 4;

QY 1739 cgaatctctgccaagcgcagctgactgccaatgataagccgaagatgtgtataggggca 1798  
 |||||  
 DB 1 CGAATCTGCGCAAGGCCGACGCTAAGCTGCAATGAGCCGACAGATTGTTATAGGGGCA 60  
 |||||  
 QY 1799 cctcgcaccaagcgcagacacattttaaagaataacatctctcctcagccacgtaccccat 1858  
 |||||  
 DB 61 CCGCGGCCAAGCGGAGACCATTTTAAAGAAATTAACATCTCTTCAGGCGACGTAACCCCATG 120  
 |||||  
 QY 1859 gacctctacagagacccctctgagcaactggaactatcttccagaagtcaggagattccagg 1918  
 |||||  
 DB 121 GACCTCTCAAGAGACCCCTCTGAGCAACTGACATCTTTCCAGAGTCCAGGATTTCCAGG 180  
 |||||  
 QY 1919 ttgaatacaagaagctccccaacaaacaaagaagaagacttctatctctataactgct 1978  
 |||||  
 DB 181 TTAATATACAAAGACTTCCCAAAAACACAGAACGAATTTGTATCTCTTATCAATTGCT 240  
 |||||  
 QY 1979 cctctcagccacctgatacgaatgtaatcggcaagatgtgtgagatccctgccaatgata 2038  
 |||||  
 DB 241 CCGTCTCAGCACCTCTGATGATGACCATGATGCGCAAGATGATGAGATCTGCGCATGATA 300  
 |||||  
 QY 2039 tggctgagctggaacatcttaagattgtctgtcgtgattgtgacacaagaatgacagagatgc 2098  
 |||||  
 DB 301 TGGCTGGCTGGAACATCTTAAAGTTGCTGTGATGATGAGTTCAGCAAAAGATGACAGATGC 360  
 |||||

Query Match	Best Local Similarity	Score	DB	Length	752:
Matches	684: Conservative	0.0	Mismatches	38:	Indels
Query Match	17.6%	Score 617.2:	DB 137:	Length 752:	
Best Local Similarity	94.2%:	Prod. No. 1.6e-142:			
Matches	684: Conservative	0.0: Mismatches	38:	Indels	4:
				Gaps	4:

QY	1247	aatatagccgagggatcaatccgatctgcgacatcgcccgatccagcagcgcaaaaaag	1306
Db	1	AATATGGCCAGGGGATCAATCCGATTACCCACATCGCCCAATCTACACAGCAAAAAAGG	60
QY	1307	agaagaagccaaagtaacacgcctcccaacagacgagcgagccctccgcgcgcagagagtttg	1366
Db	61	AGAAGGACCCAGATACACGCTCCTCACAGAGCGAGGCTCTCCGCGCGCAGGAGTTTG	120
QY	1367	tgaatcaggtgaaggttggaacaacacactgcagagaaggaaacgagcaccacaagaagttg	1426
Db	121	TGATCAGATGTGAAGGTTTGAAACAACAACACTGCAGAAAGAACGGCAGCCAAAGAGGTGG	180
QY	1427	ccaagccgaatgcagccgagaacatgctcgagatcccttggttttaagttccgcagcggc	1486
Db	181	CCAAACGGCAATGCACCCAGAAACATGCTGGAGATCTTTGGTTTCAAGTCCCGACGGCGC	240
QY	1487	agcccaaccaaacccgactccaagtcagagaggaagaaacccataaagaacacagagagatg	1546
Db	241	AGCCCAACCAACCCCGCACTCAAGTCAAGGAGGAAGACCCATTAACAAACCA - GGGATG	299
QY	1547	gaagaagaataaccttttgaacctggtcctgggagatgaaatctggactagttaataaag	1606
Db	300	GAGAAAGATAAACCTTTTGTGAACCTGGCTCTGGGGATGAAATAAGGACATGATTAAG	359
QY	1607	aggaatgaatcaagatgctctatctaaagtcataagcagcgtgcgtcgtaattctcca	1666
Db	360	AGGATGAGTTCAGGATGGCTTATCTATCACTCAATGCACAGCTGCTGCTGGAAATCTTCCCA	419
QY	1667	tgtgtcccgaggtgcgcccagcgtctagaagtttagtcaagaacatcacacaagaattta	1726
Db	420	TGTGTCCCGAGGTGCGCCAGGCTGTAGAGATGTACTCAAGGACATCACACCAAGATTTTA	479
QY	1727	ccaaggcagcttcggaatctctgcgaagccacggttaactgcatagatagcccagagttgt	1786
Db	480	CCAGGACAGCTCCGATCTCTCCCAAGGCGACGGTAACTGCCATGATAGCCGAGAGTTGT	539
QY	1787	tgtatgggggcaacctcgccacacagccgagaccattttaa - gaataacatctcttcagc	1845
Db	540	AGTATGGGGGCACTCGCCCAACGCCAGAGATCATTTTAAAGAATTAATCATCTCTTCAGGC	599
QY	1846	caagtaaccatlgacatctctcaagagaccctctgagcaactgtagactatcttccagagtc	1905
Db	600	CACGTATCCCATANGA - CTCTACAGGATCTCTGAGCAA - TGGACTATCTTCCAGAGTTC	657
QY	1906	cagggaatccagagttggaatacaagaacttcccccaaaaaacaagaacggaatttgtatct	1965
Db	658	CAGGATCTCGGGTGATCAAAAGACTTCCCAAAAAAACAGAAATGTTGTTCTCATCA	717
QY	1966	ctatc 1971	
Db	718	ATGCTC 723	
RESULT	6		
LOCUS	BE379687	821 bp	EST
DEFINITION	60115925771 NIH_MGC_53	Homo sapiens cDNA clone IMAGE:3510806	3',
ACCESSION	BE379687		
VERSION	BE379687.1	GI:9325052	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:		
AUTHORS	Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.		
TITLE	NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Tel.: (301) 496-1550		
	Email: Robert_Strausberg@nih.gov		



```
Oy 2673 agtaccacacatcacaccactgcttctctcccaacagtgatcgttattcattcat 2732
    |||
Db 121 agtaccacacatcacaccactgcttctctcccaacagtgatcgttattcattcat 180
Oy 2733 tattctcttgatctgatatgacactatataaaatttcatttggaattctcactcgt 2792
    |||
Db 181 TATTTCCTTTGATGATATGACACTATATAAATTTTCATTGGAATTTCTCAATTGT 240
Oy 2793 atctcgttaataagcacagcttggaactctgcgagactgacttaccataatcag 2852
    |||
Db 241 ATCTGTTAAATAGCAGCATTTGGAACCTGCTGAGACTGACTTATCAATAAATCTAAC 300
Oy 2853 cgacaagaatcatcatcgtatgctgctgctgctgctgctgctgctgctgctgctg 2912
    |||
Db 301 CGACAAATCATATTCATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 360
Oy 2913 gacagcttcagtgatgacaaattgtgctcctcctgctgctgctgacgaaacagctc 2972
    |||
Db 361 GACAGTTTCAGTGAATGCAAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
Oy 2973 caaaacacttgatagctcctccacagctgctgctgctgctgctgctgctgctgct 3032
    |||
Db 421 CAAAAACCTTGAAATAGCTCCACAGTGTATAAATGGAATTTGGAATTTTAAAC 480
Oy 3033 tttaagatcatctgctgctccttcttcttcttcttcttcttcttcttcttcttct 3092
    |||
Db 481 TTTAGATATCATTTGTTGTTCCATTTTTCATTTTTCATTTTTCATTTTTCATTTTTC 540
Oy 3093 cttaaaagacttgatcctcgtctttaaagattataaaacactgctgctatcat 3152
    |||
Db 541 CTTAAATGAACCTTGATCTGCTTTTAAAGATTATTAACATTTGATCTATACAT 600
Oy 3153 atgg 3156
    |||
Db 601 ATGG 604

RESULT 8
LOCUS BE613972 686 bp mRNA EST 24-AUG-2000
DEFINITION 601504093F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3905783 5',
    mRNA sequence.
ACCESSION BE613972
VERSION BE613972.1 GI:9895569
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 686)
    NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
    Tel: (301) 496-1550
    Email: Robert.Strausberg@nih.gov
    Tissue Procurement: ATCC
    CDNA Library Preparation: Life Technologies, Inc.
    CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
    DNA Sequencing by: Incyte Genomics, Inc.
    Clone distribution: MGC clone distribution information can be
    found through the I.M.A.G.E. Consortium/LNL at:
    http://image.lnl.gov
    Plate: L1CM700 row: j column: 24
    High quality sequence stop: 675.
    Location/Qualifiers
        1..686
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:3905783"
            /clone_id="NIH_MGC_71"
            /tissue_type="leiomysarcoma"
```

```
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."
BASE COUNT 226 a 156 c 196 g 108 t
ORIGIN

Query Match 17.2% Score 603.8; DB:110; Length 686;
Best Local Similarity 97.1%; Pred. No. 3.4e-139;
Matches 669; Conservative 0; Mismatches 12; Indels 8; Gaps 5;

Oy 854 aacacgacgctgctcgaacgcttgagatccttcgaatgagccctgcgaagagc 913
    |||
Db 1 AACAGATGCTGCTGCCAAGGCTTGAGGATCTCCAGATAGCCCTGCCAGAGAGC 60
Oy 914 tggagtgatggaagagaatccgaagaagaatctcaataaattcgaataaagtcag 973
    |||
Db 61 TGAAGTGAATGGAAGAGAAATCCGAAGAAATAATCTCAATTAATTAATCTCAAG 120
Oy 974 tgttgagattgcaattgaaggaactgctgctgctgctgctgctgctgctgctgctg 1033
    |||
Db 121 TGTTCGATGATGCTTAAAGGAACTTGCTGTAATTTGAGGTGCGGGAGAGTG 180
Oy 1034 gccaccccaatggaagacttgtagcaaggttcggttgaggagttgtgaggag 1093
    |||
Db 181 GCCACCCACATGAGAAAC-TTGACCAAGATTTCGGT--GGGAGATTGTGGGGAG 237
Oy 1094 gtgaagggaagaagcaagaatttcaagaanaatgcccgaatgcttcttgaggagc 1153
    |||
Db 238 GTGAAGGGAAGAAAGCAAGAAATTTCAAGAAATGCGCCCTACTCTTCTTGAGGAGC 297
Oy 1154 tgaagaattaccgcccctgctgctgctgctgctgctgctgctgctgctgctgctgct 1213
    |||
Db 298 TGAAGAAATTACCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 357
Oy 1214 caaaacccatagtcagagccacagacagccagaaatgctgcaagggatcaatccgat 1273
    |||
Db 358 CAAACCCATAGTCAAGCCACAGACCAAGCCAGAAATATGAGCA--GGATCAATCCGATTA 416
Oy 1274 gccgactggcccaagatcagcagcgaaggaaggaagagccagagtaacgcctcc 1333
    |||
Db 417 GCCGACTGGCCAGATCCAGCAGGCAAGAAAGAGAAAGAGAGCAGATACAGCTCTCA 476
Oy 1334 cagaagcagcgtcctccgagcgcgcaaggttgtagatgagtgaggttgaagacaca 1393
    |||
Db 477 CAGAGCGAGGCTCTCCGCGCCGCGCAGGAGTTTGATCCAGGTGAGGTTGGAACCA 536
Oy 1394 ctgcagaaggaagcagcagcacaagaaggtggtccaaagcgaatgcaagcgaagacatgc 1453
    |||
Db 537 CTGCAGAAAGAAAGGGGCAACAAGAAAGGTGGCCCAACCGCAATGCAGCGAATGC 596
Oy 1454 tggagatccttgg-tttaaagtcctcagcagcgagcccaaacccgcactcaagtc 1512
    |||
Db 597 TGAAGATCTTGTGTTTCAAGGTCCCGAGGCGCAGCCACAAA---CCGACTCAAGTCA 653
Oy 1513 gagagagaagacacacataaagaagaacagg 1541
    |||
Db 654 GAGGAGAAAGACACCATTAAGCAACAGGG 682

RESULT 9
LOCUS AV070406 674 bp mRNA EST 09-OCT-2000
DEFINITION AV070406 ADB Homo sapiens cDNA clone ADBCM03 5', mRNA sequence.
ACCESSION AV070406
VERSION AV070406.1 GI:10724671
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 674)
```

**AUTHORS** Peng, Y., Song, H., Huang, Q., Huang, C., Gu, Y., Yang, Y., Gao, G., Xiao, H., Xu, X., Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R., Chen, J., Chen, Z., and Han, Z.  
**TITLE** Homo sapiens CDNA ADB clones  
**JOURNAL** Unpublished (2000)  
**COMMENT** Contact: Zeguang Han  
 Chinese National Human Genome Center at Shanghai  
 351 Guo Shouling Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
 201203, P. R. China  
 Tel: 86-21-50801919(ex.45)  
 Fax: 86-21-50801922  
 Email: hanzg@hgc.sh.cn  
 This clone is available at CHGC in Shanghai.

**FEATURES**  
 source  
 1. 674  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_id="ADB"  
 /issue\_type="Adrenal gland"  
 /dev\_stage="Adult"  
 /lab\_host="SOLR"  
 /note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2: XhoI"

**BASE COUNT** 197 a 148 c 113 g 215 t 1 others  
**ORIGIN**

Query Match 17.2%; Score 602.2; DB 38; Length 674;  
 Best Local Similarity 98.4%; Pred. No. 8.6e-139;  
 Matches 618; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

QY 2510 tctcaagctgctcactcaacacattggacaacacctgacagccatgcccagagag 2569  
 |||||||  
 Db 1 tctcaagctgctcactcaacacattggacaacacctgacagccatgcccagagag 60

QY 2570 gctcttgacccgccccagacctaagaacacagaaatcgtctcactacagc 2629  
 |||||||  
 Db 61 gcctttgacccgccccagacctaagaacacagaaatcgtctcactacagc 120

QY 2630 tgaaccaactttctagtggtgcaagccccacacctctctgagtaaccacacac 2689  
 |||||||  
 Db 121 tgaccacaactttctagtggtgcaagccccacacctctctgagtaaccacacac 180

QY 2690 cactgctttcttccaaagtgatctgtatctttagttcattatttcttcttgatga 2749  
 |||||||  
 Db 181 cactgctttcttccaaagtgatctgtatctttagttcattatttcttcttgatga 240

QY 2750 tatgacacataaaatttcatltagagaatttcacatgtatctagttaaatagac 2809  
 |||||||  
 Db 241 tatgacacataaaatttcatltagagaatttcacatgtatctagttaaatagac 300

QY 2810 agtttggaactgtctgagactgacttataaatactaacgacaagaatcatacc 2869  
 |||||||  
 Db 301 agtttggaactgtctgagactgacttataaatactaacgacaagaatcatacc 360

QY 2870 atgtgtatgtgtgttagacatttatttcatltagactaacccagagacggttcaagtgc 2929  
 |||||||  
 Db 361 atgtgtatgtgtgttagacatttatttcatltagactaacccagagacggttcaagtgc 420

QY 2930 aaattgtgctcccttgctgacgtgacaacagctcgtgacttcaaaaactttaaag 2989  
 |||||||  
 Db 421 aaattgtgctcccttgctgacgtgacaacagctcgtgacttcaaaaactttaaag 480

QY 2990 tctccacagctgtataaatttgacaatttagaactttaaactttagatgatcattgg 3049  
 |||||||  
 Db 481 tctccacagctgtataaatttgacaatttagaactttaaactttagatgatcattgg 540

QY 3050 ttcatatttatttatttatttatttatttatttatttatttatttatttatttatttatt 3109  
 |||||||  
 Db 541 ntcca-ctttatttatttatttatttatttatttatttatttatttatttatttatttatt 599

QY 3110 ctctgttttaagaattataaaaaaacat 3137  
 |||||||  
 Db 600 ctctgttttaagaattataaaaaaacat 627

**RESULT** 10  
 BE674157  
 LOCUS BE674157 616 bp mRNA EST 08-SEP-2000  
 DEFINITION 7d76a04.x1 NCI-CGAP\_Lu24 Homo sapiens CDNA clone IMAGE:3278862 3'  
 similar to TR:095793 095793 STAU6N PROTEIN. ;, mRNA sequence.  
 ACCESSION BE674157  
 VERSION BE674157.1 GI:10034698  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

**REFERENCE**  
 1 (bases 1 to 616)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
**AUTHORS** National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
**JOURNAL** Unpublished (1997)  
**COMMENT** Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1350  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LINL, send email to:  
 info@image.llnl.gov  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 469.

**FEATURES**  
 source  
 1. 616  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_id="IMAGE:3278862"  
 /clone\_lib="NCI-CGAP\_Lu24"  
 /issue\_type="cardinaloid"  
 /lab\_host="DH10B"  
 /note="Organ: Lung; Vector: pTR73D-Pac (Pharmacia) with a  
 modified polylinker; Plasmid DNA from the normalized  
 library NCI-CGAP\_Lu5 was prepared, and ss circles were  
 made in vitro. Following HAP purification, this DNA was  
 used as tracer in a subtractive hybridization reaction.  
 The driver was PCR-amplified cDNAs from a pool of 5,000  
 clones made from the same library (clonoids  
 1414920-1417991 and 1520904-1522439). Subtraction by Bento  
 Soares and M. Fatima Bonaldo."

**BASE COUNT** 179 a 153 c 139 g 141 t 4 others  
**ORIGIN**

Query Match 17.0%; Score 594.4; DB 111; Length 616;  
 Best Local Similarity 98.4%; Pred. No. 7.3e-137;  
 Matches 598; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1519 aagacaccataaagaacacagagatgagaagaataacatttctgaacctgctct 1578  
 |||||||

QY 1579 gggagatgaataagagactagtaataaagagagatgactgacttataatgcat 1638  
 |||||||

QY 1639 cagcagctgctgtgtaattcttccatgtgtccgaggtgcgcagagctgtagagatt 1698  
 |||||||

QY 1699 agtcaagacatcacacaaagatttaccagggcagctccgaatctctgccaagggcag 1758  
 |||||||



```

REFERENCE      Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.
AUTHORS        1 (bases 1 to 612)
                Wu, T., Qian, B., Huang, Q., Kang, B., Gao, X., Xu, Z., Xiao, H.,
                Xu, X., Li, N., Peng, Y., Liu, F., Ou, J., Song, H., Cheng, Z., Zeng, L.,
                Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Yang
                Y., Gu, Y., Chen, Z. and Han, Z.
TITLE          Homo sapiens CDNA GK- clones
JOURNAL        Unpublished (2000)
COMMENT        Contact: Zeguang Han
                Chinese National Human Genome Center at Shanghai
                351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
                201203, P. R. China
                Tel: 86-21-50801919(ex. 45)
                Fax: 86-21-50801922
                Email: hanzg@chgc.sh.cn
                This clone is available at CHGC in Shanghai.

FEATURES
  source       Location/Qualifiers
               1..612
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone_lib="GKCF08"
               /clone_lib="GKC"
               /tissue_type="hepatocellular carcinoma"
               /dev_stage="Adult"
               /lab_host="SOLR"
               /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
               XhoI"

BASE COUNT    149 a      164 c      125 g      174 t

ORIGIN
Query Match      16.3%; Score 571.4; DB 38; Length 612;
Best Local Similarity 98.4%; Pred. No. 3.8e-131;
Matches 611; Conservative 0; Mismatches 1; Indels 9; Gaps 3;

QY 2352 ggtgacctggcggtccctcctcgaatagcagctgctgtggygaagaagggctgccagcc 2411
    |||
Db 1 GGTGACCTGGCGGTCCTCCATAGCAATAGCACTGCTGTGGGAAGAAGGCTGCCAGCC 60

QY 2412 agctggttcccggaacccagcaatcacaacccctgggaacctccggtgttgccttt 2471
    |||
Db 61 AGCTGTTCTCCCGGAGACACAGACAGATCCACACCTGGGACCTCCGTTGGTTTTT 120

QY 2472 ttcttcctctgtgtaaaagaagcagcagccctctcgaactggtcactcagac 2531
    |||
Db 121 TTTTTCCTCTGTGTAAGAAGAAAGGACGACCCCTTCTCAAGCTGGCTCACTCAGAC 180

QY 2532 acattgggaacaaacctgagcagcaltgcagaagaagggcctttgacccggcccaagact 2591
    |||
Db 181 ACATTGGGAACAAACCTGGACAGCCATGCGACAGAGAGAGGCTTTGACCGCCACAGACT 240

QY 2592 aaaaagcagagaagaataaataatgcttctactcgaagctgagcccaacttttcaagtgtc 2651
    |||
Db 241 AAAAGCAGCAGAGAAAATGTAATGTTCTCTACACGCTGACCACTTTTCTAGTGTGC 300

QY 2652 caagggcccaacacccctcctgcaatgaccacacatcaccaactcttctcctccacagt 2711
    |||
Db 301 CAGGGCCCAACACCTCTCGCACTACCCACACATCACCACTCTTCTTCTTCAACAGT 360

QY 2712 gatctgattcttagttcaattatcttcttctgattgatatgacaactataaaatttc 2771
    |||
Db 361 GATCTGATTCTTAGTTCAATTATTTCTTTGATTGATATGACACATATAT---TTTTC 416

QY 2772 atttggaaatttctcaattgatatgtaataaagaacagtttggaaactgtgtcgaagac 2831
    |||
Db 417 ATTtGGAATTTCCTCAATTGTATCTAGTTAAATGACACAGTTGGAAACTGTCCTAGAC 476

QY 2832 tgaattcacaataactcaacgacacaaagatacatccatgtgtatgtgttgaagattt 2891
    |||
Db 477 TGACTTATCAATATCTAAGCAGCAAGATCTATATCCATGTATGTGTGTGACATTTT 536

QY 2892 ctatttcattgacaaaccagagacagtttcagtgtacaaattgtgtccctctgttca 2951
    |||

```

```

Db 537 TTAATTCATTGACTAACCCAGACAGACTTTCAGTATGC--ATTGTGTGCCCTCTGTTCA 594
QY 2952 gctgaacagctcctgacctt 2972
    |||
Db 595 GCTG---CAGTCTCTGACATT 612

RESULT 13
LOCUS      BE617223
DEFINITION BE617223
            601441906f1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3846265 5',
            mRNA sequence.
ACCESSION  BE617223
VERSION    BE617223.1 GI:9888161
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.
REFERENCE  NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
            1 (bases 1 to 749)
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.llnl.gov
            Plate: LNC545 row: k column: 02
            High quality sequence stop: 705.

FEATURES
  source       Location/Qualifiers
               1..749
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="IMAGE:3846265"
               /clone_lib="NIH_MGC_65"
               /tissue_type="adenoecarcinoma"
               /lab_host="DH10B (phage-resistant)"
               /note="Organ: colon; Vector: pCMV-Sport6; Site_1: NotI;
               Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
               Average insert size 1.8 kb. Library constructed by Life
               Technologies."

BASE COUNT    244 a      166 c      223 g      116 t

ORIGIN
Query Match      16.1%; Score 565.4; DB 110; Length 749;
Best Local Similarity 95.1%; Pred. No. 1.2e-129;
Matches 714; Conservative 0; Mismatches 21; Indels 16; Gaps 12;

QY 835 aagaacagaagctcgaaacacgaatgctgctgccaac--gcgttgaagatcctgcaagaa 893
    |||
Db 1 AAGACAGACAGAGCTGCGAAGACAGATGCTGCTGCCAAAGCGCTTGAGATCTGCAGAA 60

QY 894 tgaagcccttccagagaaggtctgaagtgaaatgaaagaatccgaagaagaataatcaca 953
    |||
Db 61 TGAGCCCTCCAGAGAGAGCGGTGAGATGGAAGATTCGGAAGAAATCTCAAA 120

QY 954 taaactgaataatgaatgaatgtt--gaagctgcaactaaaggaactgtcctgtgaatt 1012
    |||
Db 121 TAAATCTGAATATGAATCAACTGTTGTGAGATTCACCTTAAACGAACTTGCCTGTGATT 180

QY 1013 tgaagtgagccggggagagatggcccaaccccaatgaagaacttggacaaaggtttcgg 1072
    |||
Db 181 TCGAGGTGGCCCGGAGAGAGGCGCCACCCACATGGAAGAAC--TTGTACCAAGTTTCGG 239

QY 1073 ctgggagtttgagggggaagtgaaagggaagaagcaagaacttcaagaagaatgacg 1132
    |||

```

```

Db      240 TTGGGAG-TTGTGGGAGGTGAAGGAAAGCAAGACAGATTTCAAAAGAAATGCCG 298
Oy      1133 ccatagcgtctttagagagcgtgaagaagttaccgcccctgcctcagttgaacgaatga 1192
Db      299 CCATAGCTGTTCTTTAGAGGCGTGAAGAGTTACCG-CCCTGCTCCTCAGTTGACGAGTAA 357
Oy      1193 aaccagatcaaaaagaagaacacacatagtcacagcagacaagccagaaatag 1252
Db      358 AGCCTAGATCAAAAAGAAACAAACCCATAGTCAGGCCACAGACAGCCAGAAATAG 417
Oy      1233 gccagggatcaatccagattagccgactgagccagatccagcaggaacaaagaagaag 1312
Db      418 GCCA-GGGATCAATCCGATTAAGCCACCTGGCCAGATCCAGCAGCAAAAAGGAGAAAG 476
Oy      1313 agccagagagacagcgtctccacagagcgagagcgtcccgccgagggaggttgatgac 1372
Db      477 AGCCAGAGACAGCGCTCCACAGAGCAGAG-CTCCCGCGCCGAGGAG-TTGTGATGC 534
Oy      1373 aagtgaaagttgaaacacacactgacagagagcagcagcaac-aagaaggtagcgaag 1431
Db      535 AGGTGAAGTGGGAAACCACTGCAGAAAGAACGGCCACCAAAAGAGGTGGCCAG 594
Oy      1432 cgcgaatgcagcagagaaatgctgtagatccttgatt--caagtcgccgacgagcagc 1489
Db      595 CGCAATGCAGCGAGAACATGCTGAGATCCTTGGTTTCAAAAGTCCCGCAGGCGCAGC 654
Oy      1490 ccacaaacccgacactcaagtcagagaggaagacacccataagaacacagggatgaga 1549
Db      655 CACAAACCGGAGCTC--AGTCAGAGGAGAGACACCCCTAA--GAAACGGGGATGAA 709
Oy      1550 gaaagtaaccttttgaacctggtctcg 1580
Db      710 GACACGTAACTTTTGGAACTGGCGCGGCG 740

```

```

RESULT 14
LOCUS      BE810330      609 bp      mRNA      EST      21-SEP-2000
DEFINITION MRA-PT0052-220500-206-f06 PT0052 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BE810330
VERSION     BE810330.1 GI:10242524
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE  1 (bases 1 to 609)
AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
            Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE    20202663
COMMENT    Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?cl=4t2=MR4-PT0052-220
            500-206-f06&t3=2000-05-22&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 10
            High quality sequence stop: 589.
            Location/Qualifiers

```

## FEATURES

Location/Qualifiers

```

source
1. .609
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="PT0052"
/dev_stage="Adult"
/note="Organ: pnet; Vector: puc18; Site:1; SmaI: Site:2;
SmaI: A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT  173 a 153 c 145 g 138 t
ORIGIN
Query Match      15.8%; Score 555; DB 136; Length 609;
Best Local Similarity 97.9%; Pred. No. 4,5e-127;
Matches 594; Conservative 0; Mismatches 10; Indels 3; Gaps 3;
Oy      1586 aaatggagactagtaataagaagatgagatgagatgcctatcctaagtcacagcag 1645
Db      6 AACATGGGAGTAAGTAAT-AAGAGATGAG-TCAGATGCTTATCTAAGTCATCAGCAGC 63
Oy      1646 tgcctgtagaatctctccatgtagccgaggtgcagcagcgtgtagagttagtaag 1705
Db      64 TGCCCTGCT-GAATTTCTTCCATGATGCGCGAGTGCCTGAGCTGTAGAGTTACTCAAG 122
Oy      1706 gacatcaaccaaagaatttaccagggcagctcgaaatctgccaagggccacgtaactg 1765
Db      123 GACATCACACCAAAAGATTTTACAGAGGACGCTCCGAATCTTCCAGGGCCAGGTAACG 182
Oy      1766 ccataatgcccgaagatgtgtgtatgtagggagcactgcgcccacacagagaccatttaa 1825
Db      183 CCATATATAGCCCGAGAGATGATGATGGGGCACTCTGGCCACAGCCGAGACATTTTAA 242
Oy      1826 agaataacatctcttcacagccacgtacccatgagactctcagacagaccctctgagcaac 1885
Db      243 AGAATAACATCTCTTCAGAGCCACGTACCCCATGAGACTCTTCAGAGACCCCTTGAGCAAC 302
Oy      1886 tggactatcttcacagagtcagggattccaggttgaatlaaagaacttcccaaaaaca 1945
Db      303 TGGACTATCTTCCAGAGCTCAGGATTCACAGTGAATACAAAGACTTCCCAAAAACA 362
Oy      1946 aaaaagaagaatttgatctctatcaaatgtctctcagcaccactctgatacagcatg 2005
Db      363 ACAAGAACGAATTTATCTTATCAATGAGCTCTCTCAGCCACCTCTGATCAGCCATG 422
Oy      2006 gtaatgcgaagatgtgagatcctcctcagatagatagtcgtcgtgtaaacatctaaagttgc 2065
Db      423 GATATGGCAGAGATGTGAGAGTCTCTCCATGATATGCTGCGCTGAACATCTTAAAGTTGC 482
Oy      2066 tgcctgagttgaccaaacaagaatagacagatgccaagaagaaaggaacagacatgctgt 2125
Db      483 TGTCTGAGTAGGAGCAACAACTACAGAGATCCAAAGAACAGAAACGACCAAAAGTCTCG 542
Oy      2126 tctgtggagagtgctggaacctttctgccaatgaacacatataaaatcccaacatata 2185
Db      543 TGTGTGGAGGTGTGAACCTTTCTGGCATGAACCATTTAATAATCTCAACATATATA 602
Oy      2186 ctgaaaa 2192
Db      603 CTGGCAA 609

```

```

RESULT 15
LOCUS      BE082712/C      615 bp      mRNA      EST      12-JUN-2000
DEFINITION RC2-BT0642-150200-012-408 BT0642 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BE082712
VERSION     BE082712.1 GI:8473017
KEYWORDS   EST.
SOURCE     human.

```



ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 615)  
 REFERENCE  
 AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H., Brunstein, A., deoliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J.  
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 MEDLINE 20202663  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?l=6t2=RC2-BT0642-150>)  
 200-012-a086t3=2000-02-15&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 75  
 High quality sequence stop: 609.  
 Location/Qualifiers  
 1. 615  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="BT0642"  
 /dev\_stage="Adult"  
 /note="Organ: Breast; Vector: puc18; Site\_1: Sma1; Site\_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
 BASE COUNT 110 a 159 c 153 g 193 t  
 ORIGIN  
 Query Match 15.7%; Score 551.6; DB 134; Length 615;  
 Best Local Similarity 97.4%; Pred. No. 3.2e-126;  
 Matches 571; Conservative 0; Mismatches 14; Indels 1; Gaps 1;  
 QY 1100 ggaagaagcaagaatttcacaaagaataccgccatagctgttcttgaaagctgaaga 1159  
 DB 615 GGAAGAAGCAAGAAATTTCAAGAAAAATGCCCATAGCTGTCTTGAGAGAGCTGAAGA 556  
 QY 1160 agttaccgccccctgctgcagtgtaacgaagtaagcctagaatcaaaaaaacaac 1219  
 DB 555 AGTTACGCCCCCTGCTGCGAGTAGAAGAGTAAGCCTAGATCAATAAAGAAAAACCAAC 496  
 QY 1220 ccatagtcacagcacagacagccagaaatatacgcaaggagatcaatccgattagccgac 1279  
 DB 495 CCATAGTCACAGCCACAGACCAAGCCCAAGATATGGCCAGGGAGATCAATCCGATTAGCCGAC 436  
 QY 1280 tggcccaagatccagcagcagcaaaaagagaagagcagagtaacagctcctcaagagc 1339  
 DB 435 TGGCCCAAGATCCAGCAGGCAAAAAAGAGAGAGCCAGAGTACACGCTCTCCACAGAGC 376  
 QY 1340 gaagccctcccgccgcaaggagtgtagtcaggtgaaggttggaacacacacacag 1399  
 DB 375 GAGGCTCTCCGCGCCGCGAGGAGAGTTTGATGACAGGTGAAGGTTGGAAACACACACTCAG 316  
 QY 1400 aaggaacgggacacacaagaaggtggccaagcgcaatgcagccgagacatgctggaga 1459  
 DB 315 AAGGAACGGGACACACAAGAGAGGTGGCCAAAGCGCAATGCAGCCGAGAACATGCTGGAGA 256

QY 1460 tccttgattcacaagtcocgcagcggaagcccaaccaaccgcgactcaaaatcagaagaga 1519  
 DB 255 TCCTTGATTCCAAGTCCGCCAGCGCGAGCCACCAAAACCCGCACTCAAGTCAAGAGAGA 196  
 QY 1520 agacaccataaagaagaacccagggagatggaagaagaagtaacctttttgaaacctgctctg 1579  
 DB 195 AGACACCCATAAAGAAACCAAGGAGATGGAAGAAAGTAACCTTTTGAACCTGGCTCTG 136  
 QY 1580 gggatgaaatgggactagtaataaagaagagatgaattcagatgacctatcctaagtcac 1639  
 DB 135 GGGATGAAATGGAGACTAGTAATAAAGAGATGAGTTCAGGATGCCCTTATCTAAGTCATC 76  
 QY 1640 agcagctgcctgctggaattcttc-ccatgtgcccgaagtcgccc 1684  
 DB 75 ACCATCTTCTGCTGGACTTCTTCACCATGTGCTCCGAGTTTCATCC 30

Search completed: April 5, 2001, 00:30:10  
 Job time: 53106 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2001, 08:30:05 ; Search time 701.87 Seconds  
(without alignments)  
1876.520 Million cell updates/sec

Title: US-09-316-048-5

Perfect score: 3506

Sequence: 1 acttcctgcgcggctgcggy.....taataactaaaaaaaaa 3506

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq\_36:\*

- 1: /SID6/gcgdata/geneseq/geneseqn/NA1980.DAT:\*
- 2: /SID6/gcgdata/geneseq/geneseqn/NA1981.DAT:\*
- 3: /SID6/gcgdata/geneseq/geneseqn/NA1982.DAT:\*
- 4: /SID6/gcgdata/geneseq/geneseqn/NA1983.DAT:\*
- 5: /SID6/gcgdata/geneseq/geneseqn/NA1984.DAT:\*
- 6: /SID6/gcgdata/geneseq/geneseqn/NA1985.DAT:\*
- 7: /SID6/gcgdata/geneseq/geneseqn/NA1986.DAT:\*
- 8: /SID6/gcgdata/geneseq/geneseqn/NA1987.DAT:\*
- 9: /SID6/gcgdata/geneseq/geneseqn/NA1988.DAT:\*
- 10: /SID6/gcgdata/geneseq/geneseqn/NA1989.DAT:\*
- 11: /SID6/gcgdata/geneseq/geneseqn/NA1990.DAT:\*
- 12: /SID6/gcgdata/geneseq/geneseqn/NA1991.DAT:\*
- 13: /SID6/gcgdata/geneseq/geneseqn/NA1992.DAT:\*
- 14: /SID6/gcgdata/geneseq/geneseqn/NA1993.DAT:\*
- 15: /SID6/gcgdata/geneseq/geneseqn/NA1994.DAT:\*
- 16: /SID6/gcgdata/geneseq/geneseqn/NA1995.DAT:\*
- 17: /SID6/gcgdata/geneseq/geneseqn/NA1996.DAT:\*
- 18: /SID6/gcgdata/geneseq/geneseqn/NA1997.DAT:\*
- 19: /SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:\*
- 20: /SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:\*
- 21: /SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3502.8	99.9	3506	21	293322 Human stauflen cDNA
2	2914.8	83.1	3217	21	293323 Human stauflen cDNA
3	2896	82.6	3142	21	293324 Human stauflen cDNA
4	2879	82.1	3260	21	293327 Human stauflen cDNA
5	2866	81.7	3190	20	X80786 DNA encoding human
6	1256	35.8	2859	21	293328 Mouse stauflen cDNA
7	1107.2	31.6	1127	20	X39991 Prostate cancer as
8	798.8	22.8	946	20	X40002 Prostate cancer as
9	755.6	21.6	844	20	X40000 Prostate cancer as
10	753.2	21.5	773	20	X39992 Prostate cancer as
11	551	15.7	825	20	X40001 Prostate cancer as
12	501.4	14.3	794	20	X39993 Prostate cancer as

C	13	433.8	12.4	660	21	279981 Human colon cancer
C	14	412.2	11.8	727	21	280269 Human colon cancer
C	15	369	10.5	769	20	217411 Human gene express
C	16	327.8	9.3	708	21	280703 Human colon cancer
C	17	286.6	8.2	305	16	T20075 Human gene signalu
C	18	215	6.1	300	20	217757 Human gene express
C	19	215	6.1	300	20	214226 Human gene express
C	20	210	6.0	767	20	216120 Human gene express
C	21	55	1.6	1000	21	A02484 Human colon cancer
C	22	52	1.5	12001	16	G76213 HSV L/ST region.
C	23	51.8	1.5	2188	20	Z77506 Human DNA-dependen
C	24	51.6	1.5	936	20	X56375 Human PKD1 gene.
C	25	51.2	1.5	53526	19	T94101 Human polycystic k
C	26	51.2	1.5	53577	17	T18551 Human PKD1 locus b
C	27	51.2	1.5	53577	19	T94108 Gene encoding a su
C	28	50.4	1.4	10732	21	A10594 Human gene express
C	29	49	1.4	1337	20	Z17263 Human gene express
C	30	48.8	1.4	867	21	Z98056 Human secreted pro
C	31	48.6	1.4	15899	19	V62164 HSV-2 strain SB5 C
C	32	48.6	1.4	16812	19	V62175 HSV-2 strain SB5 C
C	33	48.4	1.4	38186	20	Z32028 Human MERT1 relate
C	34	48	1.4	8670	20	Z32059 Human MERT1 relate
C	35	47.4	1.4	3198	20	X02974 Human IL-1ra BAC c
C	36	47	1.3	1235	20	Z16147 Human gene express
C	37	47	1.3	23142	21	A35158 Human adenosine re
C	38	47	1.3	30417	21	A35160 Human adenosine re
C	39	45.8	1.3	1446	20	Z16805 Human gene express
C	40	45.4	1.3	80240	20	V83940 NC-contlig derived
C	41	45.4	1.3	80595	20	V83939 Pseudotubular virus
C	42	45	1.3	4897	11	O03259 Human gene express
C	43	44.8	1.3	1102	20	Z16923 Human gene express
C	44	44.8	1.3	1725	12	O11243 PRV 28k gene. Pse
C	45	44.6	1.3	1558	12	Q10896 Encodes Xenopus Bo

#### ALIGNMENTS

RESULT	1
ID	Z93322 standard; cDNA; 3506 BP.
AC	Z93322;
XX	
DT	04-JUL-2000 (first entry)
XX	
DE	Human stauflen cDNA.
XX	
XX	Staufen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW	HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;
KW	acquired immune deficiency syndrome; protease; human; ss.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	409..2142
FT	/*tag= a
FT	/product= Staufen
XX	
XX	CA2238656-A1.
XX	
XX	22-NOV-1999.
XX	
XX	22-MAY-1998; 98CA-2238656.
XX	
XX	22-MAY-1998; 98CA-2238656.
XX	
XX	(UYMO-) UNIV MONTREAL.
XX	
PI	Duchaine T, Wickham L, Cohen EA, Luo M, Desrosiers L;
PI	Moulard AJ;
XX	
DR	WPI; 2000-246924/22.

DR P-PSDB: Y83023.  
XX Isolated mammalian or Caenorhabditis elegans staufen proteins useful  
PT for treating retroviral infections especially human immunodeficiency  
PT virus infections (HIV)  
XX  
PS Claim 4: Figure 1a; 96pp; English.  
XX  
CC Staufen is a RNA binding protein which interacts with double stranded  
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA  
CC transport and localization. Mammalian SP contain multiple ds  
CC RNA-binding domains. Chimeric proteins comprising portions of  
CC staufen with proteins with RNase or protease activity can be used to  
CC treat virus infections. The RNase or protease activity of the fusion  
CC protein prevents proper maturation of the virus. Mammalian staufen  
CC proteins seem to recognise double stranded RNA structure rather than  
CC any sequence specific position. 2 bacterially-expressed fusion  
CC proteins used in an RNA-binding assay (his/hsf (human SP) and  
CC MBP/msp (murine SP)) both proteins strongly bound double stranded  
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and  
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the  
CC treatment of RNA virus infections, especially those caused by  
CC retroviruses, in particular human immunodeficiency virus.  
XX  
SQ Sequence 3506 BP; 965 A; 875 C; 790 G; 876 T; 0 other;

Query Match 99.9%; Score 3502.8; DB 21; Length 3506;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 3504; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 actctctccggcgctgctggcgccgctgagccttcacggtttgctggcgccgctgctgct 60  
DB 1 attctctgcggcgctggcgccgctgagccttcacggtttgctggcgccgctgctgct 60  
OY 61 ctctctcgctccgctgctctcttcttgaacgcctcccccgcggcgccgctccctc 120  
DB 61 ctctctcgctccgctgctctcttcttgaacgcctcccccgcggcgccgctccctc 120  
OY 121 ctccagcgccactcgctctctctctctctctctctctctctctctctctctct 180  
DB 121 ctccagcgccactcgctctctctctctctctctctctctctctctctctctct 180  
OY 121 ctccagcgccactcgctctctctctctctctctctctctctctctctctctct 180  
DB 121 ctccagcgccactcgctctctctctctctctctctctctctctctctctctct 180  
OY 181 ctctctccctccctcgccgacccgacccgacccgacccgacccgacccgacccgac 240  
DB 181 ctctctccctccctcgccgacccgacccgacccgacccgacccgacccgacccgac 240  
OY 241 agcagccagagtttaataaccacttaacctctcagaactgaaacaaagacaactgtctc 300  
DB 241 agcagccagagtttaataaccacttaacctctcagaactgaaacaaagacaactgtctc 300  
OY 301 tggaaagccctcttttaaaaaagtagaacttagactctatacactgaaatcactg 360  
DB 301 tggaaagccctcttttaaaaaagtagaacttagactctatacactgaaatcactg 360  
OY 361 cacttaaaagctgttaacctgactgttcaacttctgttaaaagtaccatgtctcaagt 420  
DB 361 cacttaaaagctgttaacctgactgttcaacttctgttaaaagtaccatgtctcaagt 420  
OY 421 caagtgcagagtcagaaacccatctgctctctcagaggagccaaatactgaaacaagaac 480  
DB 421 caagtgcagagtcagaaacccatctgctctctcagaggagccaaatactgaaacaagaac 480  
OY 481 cagctctctctctcaagccttctgtagtattctcttactactagctctctgcccct 540  
DB 481 cagctctctctctcaagccttctgtagtattctcttactactagctctctgcccct 540  
OY 541 gaaatgcagagtagagccattcaaaactctgttaccctcgactgactatataaccacac 600  
DB 541 gaaatgcagagtagagccattcaaaactctgttaccctcgactgactatataaccacac 600  
OY 601 agtgcagctgcagaaagataaacccctactgtagaaactaatgacactgtgcagtgaact 660  
DB 601 agtgcagctgcagaaagataaacccctactgtagaaactaatgacactgtgcagtgaact 660

DB 601 agtgcagctgcagaaagataaacccctactgtagaaactaatgacactgtgcagtgaact 660  
OY 661 ggaataaaacccatgtataagcctgttaacccctactctcggatgcattccaccataac 720  
DB 661 ggaataaaacccatgtataagcctgttaacccctactctcggatgcattccaccataac 720  
OY 721 taacaacatgagagaggtgtctatcccccgaggtactcttaccactttccagttccact 780  
DB 721 taacaacatgagagaggtgtctatcccccgaggtactcttaccactttccagttccact 780  
OY 781 ttaacttcaagtggaactctctgttggaagagacgaacattatgtgcaaaagaaagaca 840  
DB 781 ttaacttcaagtggaactctctgttggaagagacgaacattatgtgcaaaagaaagaca 840  
OY 841 agacagctcgcaaacacagatctgtctccaaagaggttgagatctctcagaaatgaagccc 900  
DB 841 agacagctcgcaaacacagatctgtctccaaagaggttgagatctctcagaaatgaagccc 900  
OY 901 ctgcagagagagcttgaggttgatgaaagagatccgaagaagaataatcacaataact 960  
DB 901 ctgcagagagagcttgaggttgatgaaagagatccgaagaagaataatcacaataact 960  
OY 961 gaaataagtcagagtgcttgagatgtgcacttaacgggaactgtcctgtgaattcgaagt 1020  
DB 961 gaaataagtcagagtgcttgagatgtgcacttaacgggaactgtcctgtgaattcgaagt 1020  
OY 1021 gcccgaggagatggtgcccccacatgaagaacttgtagacaaaggtttcgttgaggag 1080  
DB 1021 gcccgaggagatggtgcccccacatgaagaacttgtagacaaaggtttcgttgaggag 1080  
OY 1081 ttgtgaggagagtgaaaggaaagcaagaagatttcaaaagaaatgtccgcataact 1140  
DB 1081 ttgtgaggagagtgaaaggaaagcaagaagatttcaaaagaaatgtccgcataact 1140  
OY 1141 gtcttgaggagctgaaagagttaccgcccctgtcctgcagttgaaacagatlaagcctaga 1200  
DB 1141 gtcttgaggagctgaaagagttaccgcccctgtcctgcagttgaaacagatlaagcctaga 1200  
OY 1201 atcaaaaagaaacaaacccatgaagccacagacacagacacagccacagaaatgagcagagg 1260  
DB 1201 atcaaaaagaaacaaacccatgaagccacagacacagacacagccacagaaatgagcagagg 1260  
OY 1261 atcaatccgattagccagatgagccagatccacagcagcgaacaaagagaaagagcagag 1320  
DB 1261 atcaatccgattagccagatgagccagatccacagcagcgaacaaagagaaagagcagag 1320  
OY 1321 taacagctctctcaacagagcagagctcccgcgccgagggagttgtgtatcaggttgaag 1380  
DB 1321 taacagctctctcaacagagcagagctcccgcgccgagggagttgtgtatcaggttgaag 1380  
OY 1381 gttggaacccacactgcagaaagaaacggcaccacaagaaggtggccaagcgcaatga 1440  
DB 1381 gttggaacccacactgcagaaagaaacggcaccacaagaaggtggccaagcgcaatga 1440  
OY 1441 gccgagaacatgtctggaatccctgtttcaaatcccgacgagcgagcccaacaaacc 1500  
DB 1441 gccgagaacatgtctggaatccctgtttcaaatcccgacgagcgagcccaacaaacc 1500  
OY 1501 gcaactcaagtcagagagaaagacacccataaagaaacacagggatgtgaagaaatlaacc 1560  
DB 1501 gcaactcaagtcagagagaaagacacccataaagaaacacagggatgtgaagaaatlaacc 1560  
OY 1561 ttttttgaaacctggtctcgaggatgaataatgtgacttaataaagagatgagttcagg 1620  
DB 1561 ttttttgaaacctggtctcgaggatgaataatgtgacttaataaagagatgagttcagg 1620  
OY 1621 atgccttatctaaatcagcagctgtcctgtggaattctcccatgtgtgccgaggtc 1680  
DB 1621 atgccttatctaaatcagcagctgtcctgtggaattctcccatgtgtgccgaggtc 1680  
OY 1681 gccagagctgtagaggttaagtcagagacacacacaaagatttaccagggcagctcgg 1740  
DB 1681 gccagagctgtagaggttaagtcagagacacacacaaagatttaccagggcagctcgg 1740

OY	1741	aatctctgcgaaggccacggttaactgcattcagatagaccgagatgtgtgtatgtggggcacc	1800
Db	1741	aatctctgcgaaggccacggttaactgcattcagatagaccgagatgtgtgtatgtggggcacc	1800
OY	1801	tcggccacagccgagagccattttaagaagaaataacatctcttcaggccgaagtaccccatgga	1860
Db	1801	tcggccacagccgagacattttaagaagaaataacatctcttcaggccgaagtaccccatgga	1860
OY	1861	cccttcacgagaccctctgtgcacactgtgacactctcttcacagatctccaggaattccaggt	1920
Db	1861	cccttcacgagaccctctgtgcacactgtgacactctcttcacagatctccagagttccaggt	1920
OY	1921	gaatatacaagaactctcccaaaaaacaacaagaagaattgtatctctatcaattgtctcc	1980
Db	1921	gaatatacaagaactctcccaaaaaacaacaagaagaattgtatctctatcaattgtctcc	1980
OY	1961	tcttagccacctctgtatctcagccatgttatccggaagatgtgtgagttctctgcacatgt	2040
Db	1961	tcttagccacctctgtatctcagccatgtgttatccggaagatgtgtgagttctctgcacatgt	2040
OY	2041	gtctgcgtctgacaactcttaagaatgttgcgtctgtgtgttgagccaacaagatagaagatgca	2100
Db	2041	gtctgcgtctgacaactcttaagaatgttgcgtctgtgtgttgagccaacaagatagaagatgca	2100
OY	2101	agaaacagaaacggaacccaatgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	2160
Db	2101	agaaacagaaacggaacccaatgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	2160
OY	2161	ccattataaataatcccaacatataatactgtgaataatactgtgaactgtcttgaataattggaat	2220
Db	2161	ccattataaataatcccaacatataatactgtgaataatactgtgaactgtcttgaataattggaat	2220
OY	2221	tctctgatcaactcccaatgt	2280
Db	2221	tctctgatcaactcccaatgt	2280
OY	2281	aagaacacagaaacacaaagagaagtcggtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	2340
Db	2281	aagaacacagaaacacaaagagaagtcggtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	2340
OY	2341	gatgtgcacactctcgt	2400
Db	2341	gatgtgcacactctcgt	2400
OY	2401	gtctgcacacagatgt	2460
Db	2401	gtctgcacacagatgt	2460
OY	2461	gtttgtgtcttttttttccctctgt	2520
Db	2461	gtttgtgtcttttttttccctctgt	2520
OY	2521	ctcacatcagacaattgt	2580
Db	2521	ctcacatcagacaattgt	2580
OY	2581	gtccacagatcttaaaagaacacagagaataatcaatgtcttctctactcagcgtgtacccaact	2640
Db	2581	gtccacagatcttaaaagaacacagagaataatcaatgtcttctctactcagcgtgtacccaact	2640
OY	2641	tctcagatgt	2700
Db	2641	tctcagatgt	2700
OY	2701	cttcccaacagtgatctgtatcttctgtatcttcaatatttttctgtgtgtgtgtgtgtgtgtgt	2760
Db	2701	cttcccaacagtgatctgtatcttctgtatcttcaatatttttctgtgtgtgtgtgtgtgtgtgt	2760
OY	2761	ataaataatttcaattgtggaatttcccaatgtatctagttaataagacagattgtggaac	2820
Db	2761	ataaataatttcaattgtggaatttcccaatgtatctagttaataagacagattgtggaac	2820

QY	2821	ttgtctcgactgacttataatctcaacccgaaaaagatcatatccatggtatggtg	2880
Db	2821	ttgtctcgactgacttataatctcaacccgaaaaagatcatatccatggtatggtg	2880
QY	2881	gttagacatttttttccatttgactaacccaagagagtttcagtgatgcaattgtgctc	2940
Db	2881	gttagacatttttttccatttgactaacccaagagagtttcagtgatgcaattgtgctc	2940
QY	2941	cctctggttcagctgaacacagctccctgactttcaaaaaccttgaatlaagtcctccacagt	3000
Db	2941	cctctggttcagctgaacacagctccctgactttcaaaaaccttgaatlaagtcctccacagt	3000
QY	3001	tgtataaatctgagacaattcttggaattttaaactcttgatgatcatcttggtccatttta	3060
Db	3001	tgtataaatctgagacaattcttggaattttaaactcttgatgatcatcttggtccatttta	3060
QY	3061	tttcaattttttttttttaaagcaacagagacttaaaagaaacttgatcctctgttttaa	3120
Db	3061	tttcaattttttttttttaaagcaacagagacttaaaagaaacttgatcctctgttttaa	3120
QY	3121	agatatttaaaaaaacatctgtgtaataataatgctctcttgaggacttaagctttcacta	3180
Db	3121	agatatttaaaaaaacatctgtgtaataataatgctctcttgaggacttaagctttcacta	3180
QY	3181	cactacagagataatgatctccaatgtagtccataataacctgcagagatattccagagtg	3240
Db	3181	cactacagagataatgatctccaatgtagtccataataacctgcagagatattccagagtg	3240
QY	3241	ctcgaatactgtttaattacatccatctaggagctgaaagaaatgaactaagtttcgtata	3300
Db	3241	ctcgaatactgtttaattacatccatctaggagctgaaagaaatgaactaagtttcgtata	3300
QY	3301	cagcgtgttgcctttccttgaatgttggttaacacgatacaagaagtggtgcactgaagctctg	3360
Db	3301	cagcgtgttgcctttccttgaatgttggttaacacgatacaagaagtggtgcactgaagctctg	3360
QY	3361	cgtgtgtgtcgtatgatggaacacctgtagccctgagagttaaagtaactgcttccattcatctg	3420
Db	3361	cgtgtgtgtcgtatgatggaacacctgtagccctgagagttaaagtaactgcttccattcatctg	3420
QY	3421	tttaagctggaattttctccccaatgtaagtaaaacttaagtgtttgtaaat	3480
Db	3421	tttaagctggaattttctccccaatgtaagtaaaacttaagtgtttgtaaat	3480
QY	3481	aaatgtaataactataaaaaaaa 3506	
Db	3481	aaatgtaataactataaaaaaaa 3506	
RESULT 2			
293323			
ID	293323 standard; cDNA; 3217 BP.		
AC	293323;		
XX	04-JUL-2000 (first entry)		
DT			
XX			
DE	Human stauflen cDNA.		
XX			
KW	Staufen: RNA binding protein; RNA/DNA hybrid; virus; retrovirus;		
KW	HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;		
KW	acquired immune deficiency syndrome; protease; human; ss.		
XX			
OS	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	363..1853	
FT		/*tag= a	
FT		/product= Staufen	
XX	CA2238656-A1.		
PD	22-NOV-1999		



Db 1212 gcaactaagtctcagagagagaagacacccataaagaacccagggtatggaagaaagttaacc 1271  
 QY 1561 ttttttgaaccctgctctggtggatgtaaaaa tgggactagttaataaagaagatgattccagg 1620  
 Db 1272 ttttttgaaccctgctctggtggatgtaaaaa tgggactagttaataaagaagatgattccagg 1331  
 QY 1621 atgacctatctaatgtaatacgaacgctgctgtctgtgaattctccatgtgtgccagagtc 1680  
 Db 1332 atgacctatctaatgtaatacgaacgctgctgtgaattctccatgtgtgccagagtc 1391  
 QY 1681 gccacagcgctgaagattgattgtaagagacatcacaccaaagattttccaggagagtc 1740  
 Db 1392 gccacagcgctgaagattgattgtaagagacatcacaccaaagattttccaggagagtc 1451  
 QY 1741 aatctccgcaagcgccacggtaactgcatgatagcccgagattgtgtgataaggggcacc 1800  
 Db 1452 aatctccgcaagcgccacggtaactgcatgatagcccgagattgtgtgataaggggcacc 1511  
 QY 1801 tcgcccacagccgagacacattttaagaataacatctctcaggccacgtaaccatgga 1860  
 Db 1512 tcgcccacagccgagacacattttaagaataacatctctcaggccacgtaaccatgga 1571  
 QY 1861 cctctcagagacccctctgagcaactgagactatcttccagagttccaggatccaggtc 1920  
 Db 1572 cctctcagagacccctctgagcaactgagactatcttccagagttccaggatccaggtc 1631  
 QY 1921 gaatacaagactctcccaaaaaacaagaagaaattgtatctcttcaattccatctcc 1980  
 Db 1632 gaatacaagactctcccaaaaaacaagaagaaattgtatctcttcaattccatctcc 1691  
 QY 1981 tctcagcaccctctgatacagacatggtatctggtgaagatgtgagatcctgtccatgata 2040  
 Db 1692 tctcagcaccctctgatacagacatggtatctggtgaagatgtgagatcctgtccatgata 1751  
 QY 2041 gctgcgctgtaacatctcttaagtgtcgtctgagttgagaccaaaagtacagagatgcca 2100  
 Db 1752 gctgcgctgtaacatctcttaagtgtcgtctgagttgagaccaaaagtacagagatgcca 1811  
 QY 2101 agaacaggaacaggaacaaatgctgtgtgtggaagg tgcgtgaacccctttctgaccatga 2160  
 Db 1812 agaacaggaacaggaacaaatgctgtgtgtggaagg tgcgtgaacccctttctgaccatga 1871  
 QY 2161 ccaatlaataaattccacaalatactatgtaaaatactgtaaaactgcttgaanaatttgaat 2220  
 Db 1872 ccaatlaataaattccacaalatactatgtaaaatactgtaaaactgcttgaanaatttgaat 1931  
 QY 2221 tctctgatacctcagtggtggtccgagagaacagtggtggaagaatgtggcagcagcagga 2280  
 Db 1932 tctctgatacctcagtggtggtccgagagaacagtggtggaagaatgtggcagcagcagga 1991  
 QY 2281 aagaacaacagaaacacaaagagagcgctgtggtcgcgctgagctgtgctgaggtttgtgt 2340  
 Db 1992 aagaacaacagaaacacaaagagagcgctgtggtcgcgctgagctgtgctgaggtttgtgt 2051  
 QY 2341 gatgagcaactcggtgagccggtggtccctacagcaatagcagctgctgtgtgggaaagaag 2400  
 Db 2052 gatgagcaactcggtgagccggtggtccctacagcaatagcagctgctgtgtgggaaagaag 2111  
 QY 2401 gctgcccagcagctgtgttctcccgagacacagcagatccacacccctgggcacctcgt 2460  
 Db 2112 gctgcccagcagctgtgttctcccgagacacagcagatccacacccctgggcacctcgt 2171  
 QY 2461 gtttggctttttttccctctgtgtgaaagaagaaacggcagacaccccttcaagctgg 2520  
 Db 2172 gtttggctttttttccctctgtgtgaaagaagaaacggcagacaccccttcaagctgg 2231  
 QY 2521 ctcaactcagacacatgtggaacaaacccctgagcagcatgcccagagagagcctttgacg 2580  
 Db 2232 ctcaactcagacacatgtggaacaaacccctgagcagcatgcccagagagagcctttgacg 2291  
 QY 2581 gccccagagctaaagaacacagagaataatcgaatgcttctactcagctgacccaactt 2640  
 Db 2292 gccccagagctaaagaacacagagaataatcgaatgcttctactcagctgacccaactt 2351

QY 2641 ttctagtgtgcccagggcccacacacctcctgtagtaaccacacatcaacacactgtcttct 2700  
 Db 2352 ttctagtgtgcccagggcccacacacctcctgtagtaaccacacatcaacacactgtcttct 2411  
 QY 2701 ctcccaacagtgatcgtatcttcttaatttcaatttcttcttgatgtatagaacactat 2760  
 Db 2412 ctcccaacagtgatcgtatcttcttaatttcaatttcttcttgatgtatagaacactat 2471  
 QY 2761 ataaattttcatcttgagaatttctcaattgtatctcagttaaaatagacagtttggaaac 2820  
 Db 2472 ataaattttcatcttgagaatttctcaattgtatctcagttaaaatagacagtttggaaac 2531  
 QY 2821 ttgtctgagactgacttlaacaaatactcaacgacaagaatataccatgtgtatgtg 2880  
 Db 2532 ttgtctgagactgacttlaacaaatactcaacgacaagaatataccatgtgtatgtg 2591  
 QY 2881 gttagacattttatctatctatctgataccacagacagtttcagtgatgcaaatgtgtgc 2940  
 Db 2592 gttagacattttatctatctatctgataccacagacagtttcagtgatgcaaatgtgtgc 2651  
 QY 2941 cctctggttcagctgtaaacgctcctgagctttcaaaaaccttgaatgaagctccacagt 3000  
 Db 2652 cctctggttcagctgtaaacgctcctgagctttcaaaaaccttgaatgaagctccacagt 2711  
 QY 3001 tgtataaatgtgacaattttagaattttaaactttagatgatatcatgtgttccatttta 3060  
 Db 2712 tgtataaatgtgacaattttagaattttaaactttagatgatatcatgtgttccatttta 2771  
 QY 3061 ttctaattttatctttgttatagtcaaacagagacttaaatgaactttgatactcgttttaa 3120  
 Db 2772 ttctaattttatctttgttatagtcaaacagagacttaaatgaactttgatactcgttttaa 2831  
 QY 3121 agattatttaaaaaaatgtgtatctataataatgctccttgaggacttaagctttcaacta 3180  
 Db 2832 agattatttaaaaaaatgtgtatctataataatgctccttgaggacttaagctttcaacta 2891  
 QY 3181 caactacagatacgtatctccatgtagtccatataaaacctgacagatgatttccagagt 3240  
 Db 2892 caactacagatacgtatctccatgtagtccatataaaacctgacagatgatttccagagt 2951  
 QY 3241 ctgcgatactgttlaattacatctccatgtaggctgtaaaagaatgacactagctttctgtata 3300  
 Db 2952 ctgcgatactgttlaattacatctccatgtaggctgtaaaagaatgacactagctttctgtata 3011  
 QY 3301 cagctgtgttctgttgaatgtgttctactgtacacagaagtggtgacactgaggtctgt 3360  
 Db 3012 cagctgtgttctgttgaatgtgttctactgtacacagaagtggtgacactgaggtctgt 3071  
 QY 3361 cgtgtgtgtccgtatgtgaaacacgtgtgagccctgaggttaagtaactgtccatctatgt 3420  
 Db 3072 cgtgtgtgtccgtatgtgaaacacgtgtgagccctgaggttaagtaactgtccatctatgt 3131  
 QY 3421 tttaacgtggaatttcttccccaatggaatgtaagtaaaacttaagtgtttgtcatcaat 3480  
 Db 3132 tttaacgtggaatttcttccccaatggaatgtaagtaaaacttaagtgtttgtcatcaat 3191  
 QY 3481 aaatggttaatacctaaaaaataaa 3506  
 Db 3192 aaatggttaatacctaaaaaataaa 3217

RESULT 3  
 Z93324  
 ID Z93324 standard; cDNA; 3142 BP.  
 XX  
 AC Z93324;  
 XX  
 DT 04-JUL-2000 (first entry)  
 XX  
 DE Human staufen cDNA.  
 XX  
 KW Staufen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;

KM HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;  
 KM acquired immune deficiency syndrome; protease; human; ss.  
 OS Homo sapiens.  
 XX  
 XX  
 XX Key Location/Qualifiers  
 FT CDS 288..1193 /tag= a  
 FT /product= Staufen protein  
 FT  
 PM CA2238656-A1.  
 XX  
 XX PD 22-NOV-1999.  
 XX  
 XX PF 22-MAY-1998; 98CA-2238656.  
 XX  
 XX PR 22-MAY-1998; 98CA-2238656.  
 XX  
 XX PA (UYMO-) UNIV MONTREAL.  
 XX  
 XX PI Duchaine T, Wickham L, Cohen EA, Luo M, Desrosiers L;  
 PI Moulard AJ;  
 DR WPI: 2000-246924/22.  
 DR P-PSDB: Y83108.  
 PT Isolated mammalian or Caenorhabditis elegans staufen proteins useful  
 PT for treating retroviral infections especially human immunodeficiency  
 PT virus infections (HIV)  
 XX  
 XX PS Claim 4; Figure 1b; 96pp; English.  
 XX  
 CC Staufen is a RNA binding protein which interacts with double stranded  
 CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA  
 CC transport and localization. Mammalian SP contain multiple ds  
 CC RNA-binding domains. Chimeric proteins comprising portions of  
 CC staufen with proteins with RNase or protease activity can be used to  
 CC treat virus infections. The RNase or protease activity of the fusion  
 CC protein prevents proper maturation of the virus. Mammalian staufen  
 CC proteins seem to recognise double stranded RNA structure rather than  
 CC any sequence specific position. 2 bacterially-expressed fusion  
 CC proteins used in an RNA-binding assay (his/hsp (human SP) and  
 CC MBP/msp (murine SP) both proteins strongly bound double stranded  
 CC RNA. Both fusion proteins also directly bound labeled ds RNAs and  
 CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the  
 CC treatment of RNA virus infections, especially those caused by  
 CC retroviruses, in particular human immunodeficiency virus. This  
 CC sequence is an alternatively spliced sequence to the one given in  
 CC 293322.  
 CC  
 CC  
 SO Sequence 3142 BP; 862 A; 778 C; 733 G; 769 T; 0 other;

Query Match 82.6%; Score 2896; DB 21; Length 3142;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2896; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 611 cagaagcataccctactgtacgaataatgcatgtgataaacttgaaaaaac 670  
 DB 247 cagaagcataccctactgtacgaataatgcatgtgataaacttgaaaaaac 306  
 QY 671 caatgtataagctgtttgaccttactctcgtatgacagtcacacataactacaatga 730  
 DB 307 caatgtataagctgtttgaccttactctcgtatgacagtcacacataactacaatga 366  
 QY 731 gagggagtgctatcccccgagtaactttaccattccagttccacacttacttacc 790  
 DB 367 gagggagtgctatcccccgagtaactttaccattccagttccacacttacttacc 426  
 QY 791 aagtgaaacttctgtgagagacagcaatgtaatgcaagaagaagacaagacgctg 850  
 DB 427 aagtgaaacttctgtgagagacagcaatgtaatgcaagaagaagacaagacgctg 486

QY 851 cgaacacagatgctgtgccaagcgtttgagatccttcagatgagccctgcagaga 910  
 DB 487 cgaacacagatgctgtgccaagcgtttgagatccttcagatgagccctgcagaga 546  
 QY 911 ggcctgaggtgataatgaaagaatccgaagaagaatctcaataatctgaataatgc 970  
 DB 547 ggcctgaggtgataatgaaagaatccgaagaagaatctcaataatctgaataatgc 606  
 QY 971 aagttttgagatttcacacttaaacgaacttgccttgtaatttcagagtgcccgagaa 1030  
 DB 607 aagttttgagatttcacacttaaacgaacttgccttgtaatttcagagtgcccgagaa 666  
 QY 1031 gtgcccacccacacatgaaagaacttgtgacaaagtttcggttgggagttgtgggg 1090  
 DB 667 gtgcccacccacacatgaaagaacttgtgacaaagtttcggttgggagttgtgggg 726  
 QY 1091 aagtgaaagggaaagcaagaagaattcaagaagaatgcccagatagctgttctgag 1150  
 DB 727 aagtgaaagggaaagcaagaagaattcaagaagaatgcccagatagctgttctgag 786  
 QY 1151 agctgaagaagttaccgcccctgctgcaggttgaaacgagtaagcctagaatacaaaaga 1210  
 DB 787 agctgaagaagttaccgcccctgctgcaggttgaaacgagtaagcctagaatacaaaaga 846  
 QY 1211 aacacaaaacccaatgtaagccacagacagccagaaatattgccaagggatccga 1270  
 DB 847 aacacaaaacccaatgtaagccacagacagccagaaatattgccaagggatccga 906  
 QY 1271 ttagccgactgtgccaagatccagcagcagcaaaaagaagaagagccagatcaagctcc 1330  
 DB 907 ttagccgactgtgccaagatccagcagcagcaaaaagaagaagagccagatcaagctcc 966  
 QY 1331 tcaacgagcgaagcctcccgcccgaggaagttgtgataagtgaaagttggaac 1390  
 DB 967 tcaacgagcgaagcctcccgcccgaggaagttgtgataagtgaaagttggaac 1026  
 QY 1391 acaactgcagaagaacgagcagcaacaaagaagttgccaagcgaatgcagccagaa 1450  
 DB 1027 acaactgcagaagaacgagcagcaacaaagaagttgccaagcgaatgcagccagaa 1086  
 QY 1451 tgcctgagatccttgttccaagtcctcgacgagcagcccaacccgactcaagt 1510  
 DB 1087 tgcctgagatccttgttccaagtcctcgacgagcagcccaacccgactcaagt 1146  
 QY 1511 cagaagagaagacacccataaagaacagggatggaagaagaatgaacttttgaac 1570  
 DB 1147 cagaagagaagacacccataaagaacagggatggaagaagaatgaacttttgaac 1206  
 QY 1571 ctgctctgggataaanaatgagactagtaataaagaagatgattcagaatgcttacc 1630  
 DB 1207 ctgctctgggataaanaatgagactagtaataaagaagatgattcagaatgcttacc 1266  
 QY 1631 taagtcatcagacgtcctgtcgtgaattctccatgctgcgaggtgcgcaagctg 1690  
 DB 1267 taagtcatcagacgtcctgtcgtgaattctccatgctgcgaggtgcgcaagctg 1326  
 QY 1691 taggagttagtaagaagatcacaccaaagaatttaccagagcagctccgaatcttgc 1750  
 DB 1327 taggagttagtaagaagatcacaccaaagaatttaccagagcagctccgaatcttgc 1386  
 QY 1751 aggcacagtaactgcataatgacccgagagttgttataaggggacaactgcgcaag 1810  
 DB 1387 aggcacagtaactgcataatgacccgagagttgttataaggggacaactgcgcaag 1446  
 QY 1811 ccgagacattttaaagaataacatctcttcagcgacgttaccatggaacttcaga 1870  
 DB 1447 ccgagacattttaaagaataacatctcttcagcgacgttaccatggaacttcaga 1506  
 QY 1871 gaccccttgagcaactggaacttctccagagtcacaggaatccaggttgatacaag 1930  
 DB 1507 gaccccttgagcaactggaacttctccagagtcacaggaatccaggttgatacaag 1566  
 QY 1931 acttcccaaaaacaaagaacagatgttcttctatacaattgctccttcagcaac 1990



```

Db 1567 attcccccaaaaacaagaacgaattgtatctcttaacatgtctctccagccac 1626
QY 1991 cctctgacgcgatgtatcgcgaagaatgtagtgcctccgacatgatatgctgcgtga 2050
Db 1627 cctctgacgcgatgtatcgcgaagaatgtagtgcctccgacatgatatgctgcgtga 1686
QY 2051 acatctaaagtgtctgtctgaattgacacaagaatcacagatgacccaagacggaa 2110
Db 1687 acatctaaagtgtctgtctgaattgacacaagaatcacagatgacccaagacggaa 1746
QY 2111 acggaccaaatgtctgtctgaattgacacaagaatcacagatgacccaagacggaa 2170
Db 1747 acggaccaaatgtctgtctgaattgacacaagaatcacagatgacccaagacggaa 1806
QY 2171 atcccaaatatatactgtaaaatctgtaaaatcttgaataatttgaaattctgtatcc 2230
Db 1807 atcccaaatatatactgtaaaatctgtaaaatcttgaataatttgaaattctgtatcc 1866
QY 2231 tccagtgaggccgagagacacggtggtgaaaggaatgtgagcagcagaggaagacacag 2290
Db 1867 tccagtgaggccgagagacacggtggtgaaaggaatgtgagcagcagaggaagacacag 1926
QY 2291 aaacacaagaagcgagctgtgagcgcgtgagctgtgaggttctgtgagcgcact 2350
Db 1927 aaacacaagaagcgagctgtgagcgcgtgagctgtgaggttctgtgagcgcact 1986
QY 2351 cggtagcctgagcggtccctacgaatagcagcgtcgtgtgaggaagaaggtgcgcagc 2410
Db 1987 cggtagcctgagcggtccctacgaatagcagcgtcgtgtgaggaagaaggtgcgcagc 2046
QY 2411 caactgtctctcccgagacacagcagatccaccccttgagccctcccggttctgtctt 2470
Db 2047 caactgtctctcccgagacacagcagatccaccccttgagccctcccggttctgtctt 2106
QY 2471 tttttccctgtgtgaaagaagaacgagcagccctctcaagcttggtcactcaga 2530
Db 2107 tttttccctgtgtgaaagaagaacgagcagccctctcaagcttggtcactcaga 2166
QY 2531 caatttgagacaacacccctgagcagcagtcgcaagaagaagccttgaacggccacagc 2590
Db 2167 caatttgagacaacacccctgagcagcagtcgcaagaagaagccttgaacggccacagc 2226
QY 2591 taaaagcaccagagaataatcaaatgtctctactcagcgtgacccaacttctatggtg 2650
Db 2227 taaaagcaccagagaataatcaaatgtctctactcagcgtgacccaacttctatggtg 2286
QY 2651 ccaagcaccacacacccctctgagtaacccacatccacactgtctctctccacaag 2710
Db 2287 ccaagcaccacacacccctctgagtaacccacatccacactgtctctctccacaag 2346
QY 2711 tgaatctgtatcttaattatctatcttcttgaatgataagcactatataaatctt 2770
Db 2347 tgaatctgtatcttaattatctatcttcttgaatgataagcactatataaatctt 2406
QY 2771 cattagagaattctcaattgtatctagttaaataagacagtttggaacttctgaga 2830
Db 2407 cattagagaattctcaattgtatctagttaaataagacagtttggaacttctgaga 2466
QY 2831 ctgaacttatcaataatctaaacgacacaagatcatatcattgtatgtgtagacat 2890
Db 2467 ctgaacttatcaataatctaaacgacacaagatcatatcattgtatgtgtagacat 2526
QY 2891 ttattatctatgactaacacagagcgttcaatgtagtgaacttggcgcctgtgtc 2950
Db 2527 ttattatctatgactaacacagagcgttcaatgtagtgaacttggcgcctgtgtc 2586
QY 2951 agctgaaacagctcctgagacttcaaaaaccttgataagctccccaagctgtataaat 3010
Db 2587 agctgaaacagctcctgagacttcaaaaaccttgataagctccccaagctgtataaat 2646
QY 3011 ggaacaattgagaatttaaaactttagatgacatgtgttccattttatctcaatttt 3070

```

```

Db 2647 ggaacaattgagaatttaaaactttagatgacatgtgttccattttatcttatttt 2706
QY 3071 attttgttaatgcaacaagacttaaatgaacttgaactctgttttaagaattattaa 3130
Db 2707 attttgttaatgcaacaagacttaaatgaacttgaactctgttttaagaattattaa 2766
QY 3131 aaaaacttgatataataataatggtctcttgagacttagcttcaactaactaagga 3190
Db 2767 aaaaacttgatataataataatggtctcttgagacttagcttcaactaactaagga 2826
QY 3191 tatgattccatgtagtcaataataacctgcagagatgatttccagagtgctgcatactg 3250
Db 2827 tatgattccatgtagtcaataataacctgcagagatgatttccagagtgctgcatactg 2886
QY 3251 ttaattacatctcatatagagctgaaagaatgacacagcttctgtatcaactgtgtt 3310
Db 2887 ttaattacatctcatatagagctgaaagaatgacacagcttctgtatcaactgtgtt 2946
QY 3311 gctttgatgtgtgttaactgtaacagagatgtgtacactgagagcctgtgtgttcc 3370
Db 2947 gctttgatgtgtgttaactgtaacagagatgtgtacactgagagcctgtgtgttcc 3006
QY 3371 gtagtgaacacctgtagccctgcagagttaagtaactgtctccatctatctgttaactgtg 3430
Db 3007 gtagtgaacacctgtagccctgcagagttaagtaactgtctccatctatctgttaactgtg 3066
QY 3431 aattttctcccatggaatgtaagtaaaacttaagtggttgcattcatataaatgtgaa 3490
Db 3067 aattttctcccatggaatgtaagtaaaacttaagtggttgcattcatataaatgtgaa 3126
QY 3491 actaaaaaataaaaaa 3506
Db 3127 actaaaaaataaaaaa 3142

```

```

RESULT 4
293327
ID 293327 standard; cDNA; 3260 BP.
XX
AC 293327;
XX
DT 04-JUL-2000 (first entry)
XX
DE Human stauflen cDNA.
XX
KW Stauflen; RNA binding protein; RNA/DNA hybrid; virus; retrovirus;
KW HIV; human Immunodeficiency virus; AIDS; RNase; fusion protein;
KW acquired immune deficiency syndrome; protease; ; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS /*tag= a
FT FT /product= Stauflen protein
FT FT
XX CA2238656-A1.
XX PN
XX PD 22-NOV-1999.
XX PF 22-MAY-1998; 98CA-2238656.
XX PR 22-MAY-1998; 98CA-2238656.
XX
XX (UYMO-) UNIV MONTREAL.
XX PA
XX PI Duchaine T, Wickham L, Cohen EA, Luo M, Desrosiers L;
XX PI Moulard AJ;
XX DR WPI: 2000-246924/22.
XX DR P-ESDB; Y83108.
XX PT

```

Isolated mammalian or Caenorhabditis elegans stauflen proteins useful

PT for treating retroviral infections especially human immunodeficiency  
PT virus infections (HIV)

PS Claim 4; Figure 1b; 96pp; English.

CC Staufen is a RNA binding protein which interacts with double stranded  
CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA  
CC transport and localization. Mammalian SP contain multiple ds  
CC RNA-binding domains. Chimeric proteins comprising portions of  
CC Staufen with proteins with RNase or protease activity can be used to  
CC treat virus infections. The RNase or protease activity of the fusion  
CC protein prevents proper maturation of the virus. Mammalian staufen  
CC proteins seem to recognize double stranded RNA structure rather than  
CC any sequence specific position. 2 bacterially-expressed fusion  
CC proteins used in an RNA-binding assay (his/SP (human SP) and  
CC MBP/SP (murine SP)) both proteins strongly bound double stranded  
CC RNA. Both fusion proteins also directly bound labeled ds RNAs and  
CC RNA/DNA hybrids. Staufen proteins may therefore be useful in the  
CC treatment of RNA virus infections, especially those caused by  
CC retroviruses, in particular human immunodeficiency virus. This  
CC sequence is an alternatively spliced sequence to the one given in  
CC 293322.

CC Sequence 3260 BP; 907 A; 795 C; 759 G; 799 T; 0 other;

Query Match 82.1%; Score 2879; DB 21; Length 3260;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2892; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

OY 612 agaaacataaacccctactgtgaactaaatgacatgacatgtgaactgtgaaaaaac 671  
DB 369 agaaacataaacccctactgtgaactaaatgacatgacatgtgaactgtgaaaaaac 428  
OY 672 aatgtataagcctgttgacccctactctcgcgagtcagtcacccataactacaatgag 731  
DB 429 aatgtataagcctgttgacccctactctcgcgagtcagtcacccataactacaatgag 485  
OY 732 aggaagctctatcccccaggtactcttaaccattccagttccaccttactatca 791  
DB 486 aggaagctctatcccccaggtactcttaaccattccagttccaccttactatca 545  
OY 792 agtggaaactctctgtgagagacagcaatttaatgcaaaagaaagacaagacagctgc 851  
DB 546 agtggaaactctctgtgagagacagcaatttaatgcaaaagaaagacaagacagctgc 605  
OY 852 gaacacagatgctgctgcacaaagcgttgagatcctctgcagatgagccctgcagagag 911  
DB 606 gaacacagatgctgctgcacaaagcgttgagatcctctgcagatgagccctgcagagag 665  
OY 912 gctggaggtgaaatgcaagaagaatccgaagaagaatcccaataaactgaaataatgca 971  
DB 666 gctggaggtgaaatgcaagaagaatccgaagaagaatcccaataaactgaaataatgca 725  
OY 972 agtctgttagatctgacttaaacggaactgctgtgaattccgaggtgagcccgagagag 1031  
DB 726 agtctgttagatctgacttaaacggaactgctgtgaattccgaggtgagcccgagagag 785  
OY 1032 tggcccaaccacatgaaagaactctgtgacaaagttccggttgaggaggtctgtgggga 1091  
DB 786 tggcccaaccacatgaaagaactctgtgacaaagttccggttgaggaggtctgtgggga 845  
OY 1092 aggtgaagagaaagcaagaagaattcaagaagaagaatgcgcgatagcgtcttgagga 1151  
DB 846 aggtgaagagaaagcaagaagaattcaagaagaagaatgcgcgatagcgtcttgagga 905  
OY 1152 gctggaagagttacgcgcctgctgcagttgaacgagtaagccttagaatcaaaaaaga 1211  
DB 906 gctggaagagttacgcgcctgctgcagttgaacgagtaagccttagaatcaaaaaaga 965  
OY 1212 aacaaacccatagtcagacagacaagccagagataatgagcagggatcaatccgat 1271  
DB 966 aacaaacccatagtcagacagacaagccagagataatgagcagggatcaatccgat 1025

OY 1272 tagccgactgcccagatccagcaagcaaaaaaggagaagacagatcacgctcct 1331  
DB 1026 tagccgactgcccagatccagcaagcaaaaaaggagaagacagatcacgctcct 1085  
OY 1332 cacaagacgaagctcccgccgcagggagttgtgatcaggtgtgaagttggaacca 1391  
DB 1086 cacaagacgaagctcccgccgcagggagttgtgatcaggtgtgaagttggaacca 1145  
OY 1392 cacttcagaagaagaaggagacacaagaagaagttgccaagcgcaatgacgcgagaacat 1451  
DB 1146 cacttcagaagaagaaggagacacaagaagaagttgccaagcgcaatgacgcgagaacat 1205  
OY 1452 gctggagatcctgtgttcaaaagtcgccagcgagcagcccaaacccgcatcaagtc 1511  
DB 1206 gctggagatcctgtgttcaaaagtcgccagcgagcagcccaaacccgcatcaagtc 1265  
OY 1512 agaggaagaagacaccataaagaacccaagggatggaagaagaatcaaccttttgaacc 1571  
DB 1266 agaggaagaagacaccataaagaacccaagggatggaagaagaatcaaccttttgaacc 1325  
OY 1572 tggctctgggagatgaaatgagactagtaaaagatgagttcagaatgcttatct 1631  
DB 1326 tggctctgggagatgaaatgagactagtaaaagatgagttcagaatgcttatct 1385  
OY 1632 aagtcacagcagctgcctgcgtggaattctccatggtgcccaggtgcgccaggtgt 1691  
DB 1386 aagtcacagcagctgcctgcgtggaattctccatggtgcccaggtgtgcgccaggtgt 1445  
OY 1692 aggaattagtcagaagacatcacacaagaatttaccagggcagctcogaatctgcca 1751  
DB 1446 aggaattagtcagaagacatcacacaagaatttaccagggcagctcogaatctgcca 1505  
OY 1752 ggcacggtgaactgcatgatagcccggaagttgttgatgggggcaacctggccacagc 1811  
DB 1506 ggcacggtgaactgcatgatagcccggaagttgttgatgggggcaacctggccacagc 1565  
OY 1812 cgagacatctttaaagaataacatctcttcagcgacgtaccccaatgaccccttcagag 1871  
DB 1566 cgagacatctttaaagaataacatctcttcagcgacgtaccccaatgaccccttcagag 1625  
OY 1872 accctctgagcaactgagatccttccagagtcagagatccaggttgaatacaaga 1931  
DB 1626 accctctgagcaactgagatccttccagagtcagagatccaggttgaatacaaga 1685  
OY 1932 ctcccccacaaacaaagaagaattgtatctcttcaattcaattgctctcctcaagccac 1991  
DB 1686 ctcccccacaaacaaagaagaattgtatctcttcaattcaattgctctcctcaagccac 1745  
OY 1992 tctgatacagcctgtatcgcgcaagaatgctgagtccttcacatgatalgtgctgcgta 2051  
DB 1746 tctgatacagcctgtatcgcgcaagaatgctgagtccttcacatgatalgtgctgcgta 1805  
OY 2052 catcttaagttgtctgtcgtgagttggaaccaagaatgacagatgccaagaaga 2111  
DB 1806 catcttaagttgtctgtcgtgagttggaaccaagaatgacagatgccaagaaga 1865  
OY 2112 cggaccaatgtctgtgtgtgaggtgtcgtgaacctttctgccaatgaaaccttataa 2171  
DB 1866 cggaccaatgtctgtgtgtgaggtgtcgtgaacctttctgccaatgaaaccttataa 1925  
OY 2172 tcccaacatataacttgaanaatacgtgaactcttgaanaatttgaattctgtatcct 2231  
DB 1926 tcccaacatataacttgaanaatacgtgaactcttgaanaatttgaattctgtatcct 1985  
OY 2232 ccagttggcccgagagacacggttggtaagaagtgtggtgacgacagaggaagaacaaga 2291  
DB 1986 ccagttggcccgagagacacggttggtaagaagtgtggtgacgacagaggaagaacaaga 2045  
OY 2292 aacaaagagagcggtctgtgcccgtgagctgtcgtgggttctgtgatggccactc 2351  
DB 2046 aacaaagagagcggtctgtgcccgtgagctgtcgtgggttctgtgatggccactc 2105

Oy	2352	gqfbaacbtgcbgltcccttcagcaatagacgcbgcbctbgbggbgaagaagagctgcaccaagcc	2411
Db	2106	gqfbaacbtgcbgltcccttcagcaatagacgcbgcbctbgbggbgaagaagagctgcaccaagcc	2165
Oy	2412	agcbgcbgtcccccgggagacaacaaacaaatccacaaccccttgagaccccgctgcttgctctt	2471
Db	2166	agcbgcbgtcccccgggagacaacaaacaaatccacaaccccttgagaccccgctgcttgctctt	2225
Oy	2472	ttttcccccgtgltgnaaagaagaacgagcaagaccctctcaagctgcbtcaactcagac	2531
Db	2226	ttttcccccgtgltgnaaagaagaacgagcaagaccctctcaagctgcbtcaactcagac	2285
Oy	2532	acattgagacaacaaocctcgagacagccatctgcacagaagagagccctttgacccggcccaagagct	2591
Db	2286	acattgagacaacaaocctcgagacagccatctgcacagaagagagccctttgacccggcccaagagct	2345
Oy	2592	aaaaagccacagagaanaataacaaatgcttccctactacagcgagcaacacttctcagtgtgc	2651
Db	2346	aaaaagccacagagaanaataacaaatgcttccctactacagcgagcaacacttctcagtgtgc	2405
Oy	2652	caagcccccaacacccctctcgtcaagltacccaacacatcaacacactgcttctctccaacagt	2711
Db	2406	caagcccccaacacccctctcgtcaagltacccaacacatcaacacactgcttctctccaacagt	2465
Oy	2712	gactctgacttctctagcttcaattttcttttgatgtgatagtaacataataaatttcc	2771
Db	2466	gactctgacttctctagcttcaattttcttttgatgtgatagtaacataataaatttcc	2525
Oy	2772	attcggaattcttccaattgtatctagtgttaataagacaggtttgaaactgtctcgagac	2831
Db	2526	attcggaattcttccaattgtatctagtgttaataagacaggtttgaaactgtctcgagac	2585
Oy	2832	tgactttatcaataatctaaocgagcaacaaagatacalccatgltgltatgtygttgaagacatt	2891
Db	2586	tgactttatcaataatctaaocgagcaacaaagatacalccatgltgltatgtygttgaagacatt	2645
Oy	2892	ttatttcatgacttaacaccccgagacagcttccagtgatgacaaatgtgtgcctcctggttca	2951
Db	2646	ttatttcatgacttaacaccccgagacagcttccagtgatgacaaatgtgtgcctcctggttca	2705
Oy	2952	gctbaaacagctgcctcgagacttccaacaaoccttgaataagctccccaagcttataaattg	3011
Db	2706	gctbaaacagctgcctcgagacttccaacaaoccttgaataagctccccaagcttataaattg	2765
Oy	3012	gacaaattaggaattttrtaaacctttagatgatacatgtgtgttcaattttatltcaatttca	3071
Db	2766	gacaaattaggaattttrtaaacctttagatgatacatgtgtgttcaattttatltcaatttca	2825
Oy	3072	tttttgttaatgcaaacagagacttaaaatgaacttgaactcctcgttttlaaagatatta	3131
Db	2826	tttttgttaatgcaaacagagacttaaaatgaacttgaactcctcgttttlaaagatatta	2885
Oy	3132	aaacattgtgatactatatacagatgcbtctcttgagagacttgaagcttcaactaacataaagat	3191
Db	2886	aaacattgtgatactatatacagatgcbtctcttgagagacttgaagcttcaactaacataaagat	2945
Oy	3192	atgtaatcccaatgtagtccatataataaactcgaagatgaatttcccaagatgctgcgataactgt	3251
Db	2946	atgtaatcccaatgtagtccatataataaactcgaagatgaatttcccaagatgctgcgataactgt	3005
Oy	3252	taattacaatccatctagggcgtgaaaaagaaatgaactcagttctcgtatatacagctgtgtg	3311
Db	3006	taattacaatccatctagggcgtgaaaaagaaatgaactcagttctcgtatatacagctgtgtg	3065
Oy	3312	cttttgttgtgtgttactttacacaggaagtgtgtgacatagagctcctcgcggtgtgtccg	3371
Db	3066	cttttgttgtgtgttactttacacaggaagtgtgtgacatagagctcctcgcggtgtgtccg	3125
Oy	3372	tatgaaaaaacctgtgtaacccctgcgagtttaagtaactgcttccaatctgtttaacgtgga	3431
Db	3126	tatgaaaaaacctgtgtaacccctgcgagtttaagtaactgcttccaatctgtttaacgtgga	3185
Oy	3432	atttttcccccaatggaatgtaagtaaaccttaagtgtttgttcatcaataaattgtaata	3491

Db	3186	attttcccccacggaatgaaagtaaacacttaagtcgttcgtcaccaataatggtacata	3245
.QY	3492	ctaaaaaiaaaaaaa	3506
Db	3246	ctaaaaaiaaaaaaa	3260
RESULT	5		
ID	X90786		
XX	X90786 standard; DNA; 3190 BP.		
AC	X90786;		
DT	13-JAN-2000 (first entry)		
XX			
DE	DNA encoding human stauflen protein.		
XX			
KW	Human stauflen protein; hstau; protein lysate; lung; kidney; testis;		
KM	ovary; brain; heart; telomerase RNA; complex; catalytic subunit;		
KW	telomerase assembly; maturation; transport; regulation; interaction;		
KM	inhibitory agent; antibody; immunoassay; ds.		
XX			
OS	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	345..1835	
FT		/*tag= a	
FT		/product= "Human stauflen protein"	
FT		/note= "Binds to telomerase RNA"	
FT	polyA-signal	3172..3177	
FT		/*tag= b	
XX			
PN	W09951255-A1.		
XX			
PD	14-OCT-1999.		
XX			
PF	06-APR-1999;	99WO-US07533.	
XX			
PR	06-APR-1998;	98US-0080783.	
XX			
PA	(UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.		
XX			
PI	Greider CW, Le S;		
XX			
DR	WPI; 1999-620168/53.		
XX			
DR	P-PSDB; Y26335.		
XX			
PT	Human stauflen polypeptide useful in methods for identifying telomerase		
PT	inhibitors -		
XX			
PS			
XX			
XX	Claim 13; Page 41-42; 50pp; English.		
CC			
CC	The present sequence is an isolated DNA encoding human stauflen (hstau)		
CC	protein. The hstau protein was detected in protein lysates from lung,		
CC	kidney, testis and ovary, but not in brain and heart. It binds		
CC	specifically to human telomerase RNA and can complex with the		
CC	telomerase catalytic subunit. The hstau protein may play a role in		
CC	telomerase assembly, maturation, transport and regulation. Interaction		
CC	between hstau and telomerase can be inhibited by administering inhibitory		
CC	agents to the cell. These telomerase inhibitors can be identified using		
CC	hstau protein. Antibodies generated against hstau can be used in various		
CC	immunoassays.		
XX			
XX			
SO	Sequence 3190 BP; 880 A; 788 C; 736 G; 786 T; 0 other;		
Query Match	81.7%; Score 2866; DB 20; Length 3190;		
Best Local Similarity	99.8%; Pred. No. 0;		
Matches 2880; Conservative	0; Mismatches 5; Indels 1; Gaps 1;		
QY	612	agaaagcataccctactctgtagaactaaatgcactctgcatgaaacttggaataaac	671

Db 305 agaaagcaltaaccctactctagtaactaaatgcactgycatlgaaacttgaaaaaac 364  
QY 672 aatgtataagcctgttggacccttactctggatgtgaagtcaccacttaactaaagaatgag 731  
Db 365 aactgtataagcctgttggacccttactctggatgtgaagtcaccacttaactaaagaatgag 424  
QY 732 aggaagtgctatcccccagagtaactttaccatcttcagcttccaccttactatata 791  
Db 425 aggaagtgctatcccccagagtaactttaccatcttcagcttccaccttactatata 484  
QY 732 agtggaaacttctgttggagagagcaagaatttaa tggcaaaaggaagacaagacagctgag 851  
Db 485 agtggaaacttctgttggagagagcaagaatttaa tggcaaaaggaagacaagacagctgag 544  
QY 852 gaacaagatgtctgtccaaagcgttggagatccttcgcaagatggccccctgcagagag 911  
Db 545 gaacaagatgtctgtccaaagcgttggagatccttcgcaagatggccccctgcagagag 604  
QY 912 gctggaggtgaa tggaaagaaatccgaagaagaataatccaataatctgaataatgca 971  
Db 605 gctggaggtgaa tggaaagaaatccgaagaagaataatccaataatctgaataatgca 664  
QY 972 agtgtttgagatctgcaactaaacggaaacttgcctgtgaaatttcgaagtgagccccggagag 1031  
Db 665 agtgtttgagatctgcaactaaacggaaacttgcctgtgaaatttcgaagtgagccccggagag 724  
QY 1032 tggccccaccccaatgaaagaaactttgtgaccaaagtttcgggttggggagttgttggggga 1091  
Db 724 tggccccaccccaatgaaagaaactttgtgaccaaagtttcgggttggggagttgttggggga 784  
QY 1092 aggtgaaagaaagaaagaaagaaatctcaagaanaaaatgcgcgcaatagcgtcttgagga 1151  
Db 785 aggtgaaagaaagaaagaaagaaatctcaagaanaaaatgcgcgcaatagcgtcttgagga 844  
QY 1152 gctgaaagaagttaacggccctgctgtgcagltgaaagagtaaaagcttaagaatcaaaaaagaa 1211  
Db 845 gctgaaagaagttaacggccctgctgtgcagltgaaagagtaaaagcttaagaatcaaaaaagaa 904  
QY 1212 aacaaacccatagtcagaagcccaagaacgaagcccaagaata tggccaggggataatccgat 1271  
Db 905 aacaaacccatagtcagaagcccaagaacgaagcccaagaata tggccaggggataatccgat 964  
QY 1272 tagcgaacttggcccaatctcaagcagagcaaaaaggaaggaagcagagtaacgctctct 1331  
Db 965 tagcgaacttggcccaatctcaagcagagcaaaaaggaaggaagcagagtaacgctctctct 1024  
QY 1332 cacagaagcgaagcctccgcgcgcagagagttgtgtatgtcaggttgaagtttggaaacca 1391  
Db 1025 cacagaagcgaagcctccgcgcgcagagagttgtgtatgtcaggttgaagtttggaaacca 1084  
QY 1392 cactgtcagaagaagaaagcgcgcacaaagaaggttggccaagcgcaatgtcagccggaagaact 1451  
Db 1085 cactgtcagaagaagaaagcgcgcacaaagaaggttggccaagcgcaatgtcagccggaagaact 1144  
QY 1452 gctggagaatcctgtgttcaaaagtcccgacgcgcagcagcccaaaacccgcactaaatg 1511  
Db 1145 gctggagaatcctgtgtgttcaaaagtcccgacgcgcagcagcccaaaacccgcactaaatg 1204  
QY 1512 agagaagagaaacccaataaagaacagggga tggaaagaaagttaacctttttgaaac 1571  
Db 1205 agagaagagaaacccaataaagaacagggga tggaaagaaagttaacctttttgaaac 1264  
QY 1572 tggccttggggatgaaatgtgagactaagtaagaagatgagttcagagatgctctatct 1631  
Db 1265 tggccttggggatgaaatgtgagactaagtaagaagatgagttcagagatgctctatct 1324  
QY 1632 aagtcatacagcagctgcctgtcgtgaatctctcca tggtcgaggggtcgcgcagagctgt 1691  
Db 1325 aagtcatacagcagctgcctgtcgtgaatctctcca tggtcgaggggtcgcgcagagctgt 1384  
QY 1692 aggaagtgttcaagagacatcacaccaaagattttacccaagcagctccgaatcctgtccaa 1751  
Db 1385 aggaagtgttcaagagacatcacaccaaagattttacccaagcagctccgaatcctgtccaa 1444

QY 1752 ggcacagtaactgtccatgatagcccgagagttgtgtgata tgggggacctgcgccacag 1811  
Db 1445 ggcacagtaactgtccatgatagcccgagagttgtgtgata tgggggacctgcgccacag 1504  
QY 1812 cgagaccattttaagaataacatctctcaagcagctaacccatlgagcctctcaagag 1871  
Db 1505 cgagaccattttaagaataacatctctcaagcagctaacccatlgagcctctcaagag 1564  
QY 1872 accctctgagcaacttggactatcttccagagtcgaaggtatccaggttgaataaaga 1931  
Db 1565 accctctgagcaacttggactatcttccagagtcgaaggtatccaggttgaataaaga 1624  
QY 1932 ctcccccaaaaacaaagaagaaatgtgtatctcttaacaaatgctctccagcacc 1991  
Db 1625 ctcccccaaaaacaaagaagaaatgtgtatctcttaacaaatgctctccagcacc 1684  
QY 1992 tctgatacagcca tgglatatggcaagatgtggagtcctgtcca tga taltgctgcgtgaa 2051  
Db 1685 tctgatacagcca tgglatatggcaagatgtggagtcctgtcca tga taltgctgcgtgaa 1744  
QY 2052 catcttaaaagtgtgtctgtgaagtgtgacccaagaagtacagagatgccaaagaagaa 2111  
Db 1745 catcttaaaagtgtgtctgtgaagtgtgacccaagaagtacagagatgccaaagaagaa 1804  
QY 2112 cggaccacatgtctgtgtgtgtggaggtgtctgaaacctttctgtgccaatgaaatlaaa 2171  
Db 1805 cggaccacatgtctgtgtgtgtggaggtgtctgaaacctttctgtgccaatgaaatlaaa 1864  
QY 2172 tcccaacatata tacttgaanaaatacgtgaacatgtcttgaanaatttgaatttctgatact 2231  
Db 1865 tcccaacatata tacttgaanaaatacgtgaacatgtcttgaanaatttgaatttctgatact 1924  
QY 2232 ccagtgtggcgcgaagacacaggttgggtlaaagatgttgggcagcagcagggaaagaacaa 2291  
Db 1925 ccagtgtggcgcgaagacacaggttgggtlaaagatgttgggcagcagcagggaaagaacaa 1984  
QY 2292 aacaaagaagcgcgtctgtggccg-gcgtgagactgtgctgtgggtttgttgtatgtgcca 2350  
Db 1985 aacaaagaagcgcgtctgtggccg-gcgtgagactgtgctgtgggtttgttgtatgtgcca 2044  
QY 2351 cgttgaacttgcgttccctacgcaatagcagctgtccttggagaaagggcctgcaccag 2410  
Db 2045 cgttgaacttgcgttccctacgcaatagcagctgtccttggagaaagggcctgcaccag 2104  
QY 2411 cagctgtgtctcccgggacacacagatccaaacccctgggcacccctcogtgtttgtctt 2470  
Db 2105 cagctgtgtctcccgggacacacagatccaaacccctgggcacccctcogtgtttgtctt 2164  
QY 2471 ttttttccctgtgttgaagaagaacggcagcagcccttccaaagcttggctcactcaga 2530  
Db 2165 ttttttccctgtgttgaagaagaacggcagcagcccttccaaagcttggctcactcaga 2224  
QY 2531 cacatltggacaacaccccttggacagcaltgcagagaagagccttggacgggcccaagac 2590  
Db 2225 cacatltggacaacaccccttggacagcaltgcagagaagagccttggacgggcccaagac 2284  
QY 2591 taaaagcacagagaataatcaaaatgtcttccactcaagcgttgaagcccaactttctagtgt 2650  
Db 2285 taaaagcacagagaataatcaaaatgtcttccactcaagcgttgaagcccaactttctagtgt 2344  
QY 2651 ccaagggcccaacacccctctgagtaaccaacaa tcaaccaatgtcttcttcccaag 2710  
Db 2345 ccaagggcccaacacccctctgagtaaccaacaa tcaaccaatgtcttcttcccaag 2404  
QY 2711 tgaatgtatcttgaatctcaatatttcttlttgaatgtgatalgaaactataaaatctt 2770  
Db 2405 tgaatgtatcttgaatctcaatatttcttlttgaatgtgatalgaaactataaaatctt 2464  
QY 2771 catctgaaatcttccaaatgtatataatgaatgaacaggttggaaactgtgtctgga 2830  
Db 2465 catctgaaatcttccaaatgtatataatgaatgaacaggttggaaactgtgtctgga 2524

QY 2831 ctgacttatacaataactaacgacaagatcatactcatgtgtatgtgttagacatt 2890  
 ///  
 Db 2525 ctgacttatacaataactaacgacaagatcatactcatgtgtatgtgttagacatt 2584  
 QY 2891 ttattattctgactaacaccagacagtttcagtgatgcaaatgtgtcccttgctc 2950  
 ///  
 Db 2585 ttattattctgactaacaccagacagtttcagtgatgcaaatgtgtcccttgctc 2644  
 QY 2951 agctgaaacagtcctcggaattccaacacgtgaataagatcccccagttgtataact 3010  
 ///  
 Db 2645 agctgaaacagtcctcggaattccaacacgtgaataagatcccccagttgtataact 2704  
 QY 3011 ggacaatttaggaattttaaacttagatgacatcttggttccattttatttcatcttt 3070  
 ///  
 Db 2705 ggacaatttaggaattttaaacttagatgacatcttggttccattttatttcatcttt 2764  
 QY 3071 attttgttaatgcaaacgagacttaatgaaacttgatcctctgtttttaagaatttaa 3130  
 ///  
 Db 2765 attttgttaatgcaaacgagacttaatgaaacttgatcctctgtttttaagaatttaa 2824  
 QY 3131 aaaaactgtgatactatacatatgctcttgagagacttagcttcaactacactacagga 3190  
 ///  
 Db 2825 aaaaactgtgatactatacatatgctcttgagagacttagcttcaactacactacagga 2884  
 QY 3191 tatgatactcatgtagtgcataataaactcgacagatgatttccacagatgctgcatactg 3250  
 ///  
 Db 2885 tatgatactcatgtagtgcataataaactcgacagatgatttccacagatgctgcatactg 2944  
 QY 3251 ttaataacttcccatatagagctgtaaaagaagaaactgaactgaatttcgtatacagctgtgt 3310  
 ///  
 Db 2945 ttaataacttcccatatagagctgtaaaagaagaaactgaactgaatttcgtatacagctgtgt 3004  
 QY 3311 gctttgagctgtgtgactgacacagaaagtgtgacactgagctcgcggtgtgtctc 3370  
 ///  
 Db 3005 gctttgagctgtgtgactgacacagaaagtgtgacactgagctcgcggtgtgtctc 3064  
 QY 3371 gtagtgaacacactgtagcctcgagctgaagtaactgcttccatcattgttttaacgctg 3430  
 ///  
 Db 3065 gtagtgaacacactgtagcctcgagctgaagtaactgcttccatcattgttttaacgctg 3124  
 QY 3431 aattttcccccagatgtaagttaaaaacttaagtgtttgtccatcaataatgtgtaac 3490  
 ///  
 Db 3125 aattttcccccagatgtaagttaaaaacttaagtgtttgtccatcaataatgtgtaac 3184  
 QY 3491 actaaa 3496  
 ///  
 Db 3185 acttaa 3190  
 ///  
 RESULT 6  
 293328  
 ID 293328 standard; cDNA: 2859 BP.  
 XX  
 AC 293328;  
 XX  
 DT 04-JUL-2000 (first entry)  
 XX  
 DE Mouse stauften cDNA.  
 XX  
 XX Stauften, RNA binding protein; RNA/DNA hybrid; virus; retrovirus;  
 KW HIV; human immunodeficiency virus; AIDS; RNase; fusion protein;  
 KM acquired immune deficiency syndrome; protease; mouse; ss.  
 XX  
 OS Mus musculus.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 325..1788  
 FT /\*tag= a  
 FT /product= Stauften protein  
 PN CA2238656-A1.  
 XX  
 PD 22-NOV-1999.

XX  
 XX 22-MAY-1998: 98CA-2238656.  
 PF  
 XX 22-MAY-1998: 98CA-2238656.  
 PR  
 XX (UWMO-) UNIV MONTREAL.  
 PA  
 XX Duchaine T, Wickham L, Cohen EA, Luo M, Desrosiers L;  
 PI Moulard AJ;  
 XX  
 DR WPI: 2000-246924/22.  
 DR P-PSDB: Y83024.  
 XX  
 PT Isolated mammalian or Caenorhabditis elegans stauften proteins useful  
 PT for treating retroviral infections especially human immunodeficiency  
 PT virus infections (HIV)  
 XX  
 PS Claim 4: Figure 1c: 96pp; English.  
 PS  
 CC Stauften is a RNA binding protein which interacts with double stranded  
 CC RNA and/or rough endoplasmic reticulum. It is involved in mRNA  
 CC transport and localization. Mammalian SP contain multiple ds  
 CC RNA-binding domains. Chimeric proteins comprising portions of  
 CC stauften with proteins with RNase or protease activity can be used to  
 CC treat virus infections. The RNase or protease activity of the fusion  
 CC protein prevents proper maturation of the virus. Mammalian stauften  
 CC proteins seem to recognise double stranded RNA structure rather than  
 CC any sequence specific position. 2 bacterially-expressed fusion  
 CC proteins used in an RNA-binding assay (his/hsf (human SP) and  
 CC MBP/msp (murine SP)) both proteins strongly bound double stranded  
 CC RNA. Both fusion proteins also directly bound labeled ds RNAs and  
 CC RNA/DNA hybrids. Stauften proteins may therefore be useful in the  
 CC treatment of RNA virus infections, especially those caused by  
 CC retroviruses, in particular human immunodeficiency virus.  
 CC  
 XX  
 SQ Sequence 2859 BP; 747 A; 750 C; 712 G; 650 T; 0 other;  
 ///  
 Query Match 35.8%; Score 1256; DB 21; Length 2859;  
 Best Local Similarity 86.7%; Pred. No. 0;  
 Matches 1420; Conservative 0; Mismatches 210; Indels 8; Gaps 3;  
 ///  
 QY 611 caagaacgaataccactgactgaaactaaatgacatgcatgaaactgtgaaacac 670  
 ///  
 Db 263 caagaacgaataccactgactgaaactaaatgacatgcatgaaactgtgaaacac 322  
 QY 671 caatgataagcctgttgacacttactctcgatgacagtcacactataactacaactga 730  
 ///  
 Db 323 caatgataagcctgttgacacttactctcgatgacagtcacactataactacaactga 382  
 QY 731 gaagaggtgtctatcccccgaagttacttaccattccagttccaccttactatc 790  
 ///  
 Db 383 gtgaggtgtctatcccccgaagttacttaccattccagttccaccttactatc 442  
 QY 791 aagtggaactctgttgagagacagaaacttaatgcaagaagaaagacaagcagctg 850  
 ///  
 Db 443 aagtggaactctgttgagagacagaaacttaatgcaagaagaaagacaagcagc 502  
 QY 851 cgaacacgatagtctgtcccaaaagcgttgagatactctgcagaaatgagccctgcagaga 910  
 ///  
 Db 503 tgaacacgatagtctgtcccccgtgcgttgagactctgcagagtgaaacccctgcagaaa 562  
 QY 911 gcttgagagtgatgagaagaaatcgaaagaagaaatcccaataaactcgaataagtc 970  
 ///  
 Db 563 gcttgagagtgatgagaagaaatcgaaagaagaaatcccaataaactcgaataagtc 622  
 QY 971 aagtggttgatgacttaaaacgaaactgcttgaaattcgaaggttgcccgagaga 1030  
 ///  
 Db 623 aagtggttgatgacttaaaacgaaactgcttgaaattcgaaggttgcccgagaga 602  
 QY 1031 gtggccacccacacatgaaagactttgtgacaaagtttcggttgaggagttgtgagg 1090  
 ///  
 Db 683 gtggccacccacacatgaaagactttgtgacaaagtttcggttgaggagttgtgagg 742

[illegible]

QY 2171 atccacacatatactacgaaaatracga -aacgtcttgaaaaatttgaaatttcgaac 2229  
 Db 1817 a-ccccaaacatactacgcgaaaatcacgcgagaactgccttgaaaatttgaaatactcgataa 1875  
 QY 2230 ctccagctggagccgagaga 2247  
 Db 1876 ctccagctggagccgagaca 1893

RESULT	7
X39991	
ID	X39991 standard; DNA; 1127 BP.

XX	XX
AC	X39991;

XX	02-JUL-1999	(first entry)
DT		

XX	Prostate cancer associated gene.
DE	

XX Cancer associated antigen; diagnosis; research; treatment; human;  
KW

KW	breast cancer; colon
KW	prostate cancer; ss

XX	Homo sapiens.
OS	

XX W09904265-A2.  
PN

XX  
PD' 28-JAN-1999.

XX 15-JUL-1998; 98WO-US14679  
PF

XX 22-JUN-1998; 98US-0102322  
PR

PR	17-JUL-1997;	97US-0896164
PR	10-OCT-1997;	97US-0061599

PR	10-OCT-1997;	97US-0061765
PR	10-OCT-1997;	97US-0948705

PR 11-OCT-1997; 97GB-0021697  
XX

PA (LUDWIG INST CANCER I  
XX

PI Chen Y, Gout I, Gure A, C  
PI Pfreundschuh M, Sahin U, s

PI	Turecl O;
XX	1000 100440.43
XX	

DR WPL; 1999-132448/11.  
XX

PT New isolated cancer associated sera from cancer patients

for the diagnosis, monitoring

PS Claim 6; page 625-626; 7/8/1998  
XX

CC The invention relates to a human

biological sample isolated from

CC binds to the NAM, an expressed product complexed with an H1N1

CC interaction between the agent  
CC determination of the disorder  
CC the disorder's membership

the diagnosis, monitoring, and characterised by the expression of the identified marker must

CC The invention provides nucleic acid sequences which are cancer associated

CC cancer, renal cancer, colon  
CC lung cancer.

Sequence 1127 BP; 349 A; 265

Continued on next page

Query Match	31.08;
Best Local Similarity	99.48;
Watchdog 1130: Configuration	

matches 1120; conservative

PS Claim 67, Page 625-626; 787pp; English.

Query Match	31.6%	Score 1107.2;	DB 20;	Length 1127;
Best Local Similarly	99.4%	Pred. No. 7.9e-274;		
Matches 1120; Conservative	0;	Mismatches 6;	Indels 1;	Gaps 1;

QY	809	gaggaacagcaatttaatgycgaaagagaaacaaagacaggtctgcgaacacagatctgctg	866
Dp	1	gaggaacagcaatttaatgycgaaagagaaacaaagacaggtctgcgaacacagatctgctg	60
QY	869	ccaaagagtttgagatctctgcgaataatgagccctctccagagaggtctggaagtgaatgaa	928
Dp	61	ccaaagagtttgagatctctgcgaataatgagccctctccagagaggtctggaagtgaatgaa	120
QY	929	gagaatctcgaaagaaagaaatctcaataaactcgaataaagtaaaagtcttgatctgcac	988
Dp	121	gagatctcgaaagaaagaaatctcaataaactcgaataaagtaaaagtcttgatctgcac	180
QY	989	ttaaagaggaactgcgtctgtaattctcgaggtgcgccgagagatgcccaccacatgta	1048
Dp	181	ttaaagaggaactgcgtctgtaattctcgaggtgcgccgagagatgcccaccacatgta	240
QY	1049	agaaacttggaccaaggttctggtgtggggagtttgggtgggaaggtgaagggaaagca	1108
Dp	241	agaaacttggaccaaggttctggtgtggggagtttgggtgggaaggtgaagggaaagca	300
QY	1109	agaaagattcaaaagaaaaatgycgcacatgcgtctctctgaagagcttgaaagatctaccg	1168
Dp	301	agaaagattcaaaagaaaaatgycgcacatgcgtctctctgaagagcttgaaagatctaccg	360
QY	1169	ccctgcgtcagttgaaacgagatlaaagcgcctgaatactaaaagaaacaaacccatgta	1228
Dp	361	ccctgcgtcagttgaaacgagatlaaagcgcctgaatactaaaagaaacaaacccatgta	420
QY	1229	agccacaagaaagcccagaataatgcccaggggatacctaactgataccgactgtgccaga	1288
Dp	421	agccacaagaaagcccagaataatgcccaggggatacctaactgataccgactgtgccaga	480
QY	1289	tcacgcagcgcaaaaagagaaagagacacagagatacagctctctcaaaagagagccctc	1348
Dp	481	tcacgcagcgcaaaaagagaaagagacacagagatacagctctctcaaaagagagccctc	540
QY	1349	cgcgccgcagggagtttggatgatacaggtgaaggtctgaaacacacactctcagaagaaagc	1408
Dp	541	cgcgccgcagggagtttggatgatacaggtgaaggtctgaaacacacactctcagaagaaagc	600
QY	1409	gcaccaacaagaaagtgcgccaagcgaatgcagccgagaaacatagtctgtagatctctggt	1468
Dp	601	gcaccaacaagaaagtgcgccaagcgaatgcagccgagaaacatagtctgtagatctctggt	660
QY	1469	tcaaaagtcccgacgcgcgcgccccaccaaacccgcactcaagtccaagagaaagaaaccca	1528
Dp	661	tcaaaagtcccgacgcgcgcgccccaccaaacccgcactcaagtccaagagaaagaaaccca	720
QY	1529	taaagaaacacagggaggtgagaaagaaagttaacctttttaacactgtgcctctgggagtgaa	1588
Dp	721	taaagaaacacagggaggtgagaaagaaagttaacctttttgaaacctgtgcctgggagtgaa	780
QY	1589	atggagactagtaataaagagagatgcagatcgaatgacctataatgaatcatcagcagctgc	1648
Dp	781	atggagactagtaataaagagagatgcagatcgaatgacctataatgaatcatcagcagctgc	840
QY	1649	ctgcgtggaattcttcccatagtgctcccgaggtgcgccagagctgttaaggttgatcaagac	1708
Dp	841	ctgcgtggaattcttcccatagtgctcccgaggtgcgccagagctgttaaggttgatcaagac	900
QY	1709	atcacacacaaagattttacacagagcgcctcgatctctgcgaagccagagctgaactgtgca	1768
Dp	901	atcacacacaaagattttacacagagcgcctcgatctctgcgaagccagagctgaactgtgca	960
QY	1769	tgatagcccgagagttgtgtatgtggggcactctgcaccaag-ccgagaccaattttaag	1827
Dp	961	tgatagcccgagagttgtgtatgtggggcactctgcaccaag-ccgagaccaattttaag	1020
QY	1828	aataaactcttcaagccagctacacccaatgagacctctcaagagacacctctgagaaactg	1887
Dp	1021	aataaactcttcaagccagctacacccaatgagacctctcaagagacacctctgagaaactg	1080

```

OY      1888 gactatcttccagagtccaggagattccggttgataacaagacct 1934
          |||
Db      1081 gactatcttncagagttccaggagattncaggttgataaccagaacct 1127
          |||

RESULT      8
X40002
ID        X40002 standard; DNA; 946 BP.
XX
AC        X40002;
XX
DT        02-JUL-1999 (first entry)
DE        Prostate cancer associated gene.
KW        Cancer associated antigen; diagnosis; research; treatment; human;
           breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
           prostate cancer; ss.
XX
OS        Homo sapiens.
XX
PN        WO9904265-A2.
XX
PD        28-JAN-1999.
PE        15-JUL-1998; 98WO-US14679.
PR        22-JUN-1998; 98US-0102322.
PR        17-JUL-1997; 97US-0896164.
PR        10-OCT-1997; 97US-0061599.
PR        10-OCT-1997; 97US-0061765.
PR        10-OCT-1997; 97US-0948705.
PR        11-OCT-1997; 97GB-0021697.
XX
PA        (LDDM-) LDDMIG INST CANCER RES.
XX
PI        Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI        pfriedrundschnuh M, Sahlin U, Scanlan MJ, Stockert E;
PI        Tureci O;
XX
DR        WPI: 1999-132448/11.
XX
PT        New isolated cancer associated nucleic acids and polypeptides -
PT        isolated using sera from cancer patients, used to develop products
PT        for the diagnosis, monitoring or treatment of cancers
XX
PS        Claim 67; Page 629-630; 787pp; English.
XX
CC        The invention relates to a method for diagnosing a disorder characterised
CC        by expression of a human cancer associated antigen precursor coded for by
CC        a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC        biological sample isolated from a subject with an agent that specifically
CC        binds to the NAM, an expression product or a fragment of an expression
CC        product complexed with an HLA molecule; and (b) determining the
CC        interaction between the agent and the NAM or the expression product as a
CC        determination of the disorder. The products and methods can be used in
CC        the diagnosis, monitoring, research, or treatment of conditions
CC        characterised by the expression of various cancer associated antigens.
CC        The invention provides nucleic acid sequences and encoded polypeptides
CC        which are cancer associated antigen precursors expressed in human breast
CC        cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC        lung cancer.
XX
SQ        Sequence 946 BP; 292 A; 208 C; 259 G; 170 T; 17 other;

Query Match .                22.8%; Score 798.8; DB 20; Length 946;
Best Local Similarity       93.6%; Pred. No. 7.4e-195;
Matches 888; Conservative   0; Mismatches 54; Indels 7; Gaps 6

OY      809 gagagacgaattaatggcnaaaagaaaagacaagacgctgcgnaaacgatgctgctg 868
          |||
Db      1 ggagcaagcaatttaattggcnaaaagaaaagacaagacgctgcgnaaacgatgctgctg 60

```

```

OY 869 ccaaaagcttgagatcctgcagaatgagccctgcagaagagctgagtgatgaa 928
DB 61 ccaaaagcttgagatcctgcagaatgagccctgcagaagagctgagtgatgaa 120
OY 929 gagaatccgaagaagaaatctcaataaactgaataaagtaagtgattgagcac 988
DB 121 gagaatccgaagaagaaatctcaataaactgaataaagtaagtgattgagcac 180
OY 989 ttaaacggaacttgctgagatctcgaggtgcccggagagtgccacccacatga 1048
DB 181 ttaaacggaacttgctgagatctcgaggtgcccggagagtgccacccacatga 240
OY 1049 agaaacttgagcacaaggttcgtgtggtggaattgtgtggtggaaggtgaagaaagca 1108
DB 241 agaaacttgagcacaaggttcgtgtggtggaattgtgtggtggaaggtgaagaaagca 300
OY 1109 agaaatctcaagaagaaatgcgcgcataagctgtctcttgaggagctgaagaattacgcg 1168
DB 301 agaaatctcaagaagaaatgcgcgcataagctgtctcttgaggagctgaagaattacgcg 360
OY 1169 ccctgcctgcagttgaaagcagaaagcctgaatacaaaaagaaacaaacccatagta 1228
DB 361 ccctgcctgcagttgaaagcagaaagcctgaatacaaaaagaaacaaacccatagta 420
OY 1229 agccacagacagccacagataatgcccagggatcgaatccgaattagccagtgccaga 1288
DB 421 agccacagacagccacagataatgcccagggatcgaatccgaattagccagtgccaga 480
OY 1289 tccagcagcacaagaaagagagagcagagtaacagctctcctcaagaagcagagcctcc 1348
DB 481 tccagcagcacaagaaagagagagcagagtaacagctctcctcaagaagcagagcctcc 540
OY 1349 cgcgcgcagcagaggttgatgagtgagaggttggaagcttggaacacacac-tgcagaagaagc 1407
DB 541 cgcgcgcagcagaggttgatgagtgagaggttggaagcttggaacacacac-tgcagaagaagc 600
OY 1408 ggcacacacaaagaaaggtgcccagcgaatgagcagagcaaatctctggaatctctgtgt 1467
DB 601 ggcacacacaaagaaaggtgcccagcgaatgagcagagcaaatctctggaatctctgtgt 660
OY 1468 ttc-aaatcccgagcagcagccacacacacacacacacacacacacacacacacacacac 1526
DB 661 ttc-aaatcccgagcagcagccacacacacacacacacacacacacacacacacacacacac 720
OY 1527 cataaagaaacagggagatgagaaagaaataacatttttgaacctggtctggggatga 1586
DB 721 attaagaaacacagggagatgagaaagaaataac--ntttgaacctggtctggggatga 778
OY 1587 aaatggagactgaataaagagatgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 1646
DB 779 aaatggag-ctgtantaaagagatgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 836
OY 1647 gctgtcctggaatctctccatggtgcccagagtcgcgccagagctgtagagtgatgaag 1706
DB 837 tgcctcctggaatctctccatggtgcccagagtcgcgccagagctgtagagtgatgaag 895
OY 1707 acatacacaagaagatttaccagggcagcttcggaatcctgcgaagggc 1755
DB 896 cccttcacccaagaatttcccggcagcttcccaatctgcgaagggc 944

```

```

RESULT 9
X40000
ID X40000 standard; DNA; 844 BP.
AC X40000;
DT 02-JUL-1999 (first entry)
XX Prostate cancer associated gene.
DE Cancer associated antigen; diagnosis; research; treatment; human;
KW

```

```

KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KM prostate cancer; ss.
XX Homo sapiens.
OS WO9904265-A2.
PN 28-JAN-1999.
PD 15-JUL-1998; 98WO-US14679.
XX PF 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0836164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 11-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX (LUDWIG INST CANCER RES.
PA Chen Y, Gout I, Gure A, O'Hare M, Ohta Y, Old LJ;
XX Pfeundschnuh M, Sahlin U, Scanlan MJ, Stockert E;
XX Tureci O;
DR WPI: 1999-132448/11.
XX New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX Claim 67; Page 629; 787pp; English.
XX The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
SQ Sequence 844 BP; 251 A; 196 C; 234 G; 156 T; 7 other:

```

```

Query Match 21.6%; Score 755.6; DB 20; Length 844;
Best Local Similarity 96.3%; Pred. No. 8.1e-184;
Matches 812; Conservative 0; Mismatches 26; Indels 5; Gaps 4;
OY 971 aagtggttgagatgagcttaaacggaacttgctgtaatttcgaggtgcccggagga 1030
DB 1 aagtggttgagatgagcttaaacggaacttgctgtaatttcgaggtgcccggagga 60
OY 1031 gtggccacccacacatgaagaacttgygaccaaagtttcgtgtggagtttggtggg 1090
DB 61 gtggccacccacacatgaagaacttgygaccaaagtttcgtgtggagtttggtggg 120
OY 1091 aagtggaagggaagaagcaagaatttcaagaataatgcccacataagctgtcttgagg 1150
DB 121 aagtggaagggaagaagcaagaatttcaagaataatgcccacataagctgtcttgagg 180
OY 1151 agctgaagaagttacggccctgctgcaagtgaacaggtlaaagcctgaatcaaaaaga 1210
DB 181 agctgaagaagttacggccctgctgcaagtgaacaggtlaaagcctgaatcaaaaaga 240
OY 1211 aaacaaaaccatagttcaagccacagacagagccagaaatagtcaggagatcaatccga 1270
DB 241 aaacaaaaccatagttcaagccacagacagagccagaaatagtcaggagatcaatccga 300

```



QY 1271 ttacccgactggtccagatccagcaggaaggaagagccagagttacacgtcc 1330  
 |||||  
 Db 301 ttacccgactggtccagatccagcaggaaggaagagccagagttacacgtcc 360  
 QY 1331 tcacagagcagagcgtcccgccgaggaaggttctgagtcaggtgaagacc 1390  
 |||||  
 Db 361 tcacagagcagagcgtcccgccgaggaaggttctgagtcaggtgaagacc 420  
 QY 1391 acacitcagaggaagcaggaaggaaggaaggttgcacagcaggtgcaggaaga 1450  
 |||||  
 Db 421 acacitcagaggaagcaggaaggaaggaaggttgcacagcaggtgcaggaaga 480  
 QY 1451 tctgtgagatccttcttcaagtcctcgagcagccacacacccgacacagt 1510  
 |||||  
 Db 481 tctgtgagatccttcttcaagtcctcgagcagccacacacccgacacagt 540  
 QY 1511 cagaagagagacacccataaagaacacagggatggaagaaagtaaccttttgaac 1570  
 |||||  
 Db 541 cagaagagagacacccataaagaacacagggatggaagaaagtaaccttttgaac 600  
 QY 1571 ctgtccttggggatggaagaaaggaagcagtaataagaaggttgcaggtcctatc 1630  
 |||||  
 Db 601 ctgtccttggggatggaagaaaggaagcagtaataagaaggttgcaggtcctatc 660  
 QY 1631 taagtcacagcagcgtcgtcgtggaattctccatgtgtgcccgaaggtccgagctg 1690  
 |||||  
 Db 661 taagtcacagcagcgtcgtcgtggaattctccatgtgtgcccgaaggtccgagctg 1720  
 QY 1691 taagg-agttaagtaagacatcacaccagaagatttaccagggcagctccgaatcctgccc 1749  
 |||||  
 Db 721 taagg-agttaagtaagacatcacaccagaagatttaccagggcagctccgaatcctgccc 779  
 QY 1750 aagggcaggttaactgcatgataagcccgagagtttctgatactggggcagctccgacaa 1809  
 |||||  
 Db 780 aagggcaggttaactgcatgataagcccgagagtttctgatactggggcagctccgacaa 836  
 QY 1810 gcc 1812  
 |||  
 Db 837 gcc 839  
 |||  
 RESULT 10  
 X39992  
 ID X39992 standard; DNA; 773 BP.  
 XX  
 AC X39992;  
 XX  
 DT 02-JUL-1999 (first entry)  
 XX  
 DE Prostate cancer associated gene.  
 XX  
 KW Cancer associated antigen; diagnosis; research; treatment; human;  
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
 KW prostate cancer; ss.  
 XX  
 OS Homo sapiens.  
 OS  
 PN MO9904265-A2.  
 XX  
 PD 28-JAN-1999.  
 XX  
 PF 15-JUL-1998; 98MO-US14679.  
 XX  
 PR 22-JUN-1998; 98US-0102322.  
 PR 17-JUL-1997; 97US-0896164.  
 PR 10-OCT-1997; 97US-0061599.  
 PR 10-OCT-1997; 97US-0061765.  
 PR 10-OCT-1997; 97US-0948705.  
 PR 11-OCT-1997; 97GB-0021697.  
 XX  
 PA (LUDW-) LUDWIG INST CANCER RES.  
 XX

PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;  
 PI Pfeundschuh M, Sahin U, Scanlan MJ, Stockert E;  
 PI Tureci O;  
 XX  
 DR WPI; 1999-132448/11.  
 XX  
 XX New isolated cancer associated nucleic acids and polypeptides -  
 PT isolated using sera from cancer patients, used to develop products  
 PT for the diagnosis, monitoring or treatment of cancers  
 XX  
 PS Claim 67; Page 626; 787pp; English.  
 XX  
 CC The invention relates to a method for diagnosing a disorder characterised  
 CC by expression of a human cancer associated antigen precursor coded for by  
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
 CC biological sample isolated from a subject with an agent that specifically  
 CC binds to the NAM, an expression product or a fragment of an expression  
 CC product complexed with an HLA molecule; and (b) determining the  
 CC interaction between the agent and the NAM or the expression product as a  
 CC determination of the disorder. The products and methods can be used in  
 CC the diagnosis, monitoring, research, or treatment of conditions  
 CC characterised by the expression of various cancer associated antigens.  
 CC The invention provides nucleic acid sequences and encoded polypeptides  
 CC which are cancer associated antigen precursors expressed in human breast  
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
 CC lung cancer.  
 XX  
 SQ Sequence 773 BP; 233 A; 204 C; 199 G; 134 T; 3 other;

Query Match 21.5%; Score 753.2; DB 20; Length 773;  
 Best Local Similarity 99.1%; Pred. No. 3.2e-183;  
 Matches 766; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 1163 taccgccctgtcctcgtgaagtaagcctagatcaaaagaacaaaccca 1222  
 |||||  
 Db 1 taccgccctgtcctcgtgaagtaagcctagatcaaaagaacaaaccca 60  
 QY 1223 tagtcaagccacagacagccagaatattgccaagggatccatccgattagccgactgg 1282  
 |||||  
 Db 61 tagtcaagccacagacagccagaatattgccaagggatccatccgattagccgactgg 120  
 QY 1283 cccagatccagcagcagcaaaagaaggaagacccagatcacgcctctcacagagag 1342  
 |||||  
 Db 121 cccagatccagcagcagcaaaagaaggaaggaagcagatcacgcctctcacagagag 180  
 QY 1343 gctcccgccgcagagagttgtgatacaggtgaagttggaacacacacacacagag 1402  
 |||||  
 Db 181 gctcccgccgcagagagttgtgatacaggtgaagttggaacacacacacacagag 240  
 QY 1403 gaacggcaccacaagaagaagtgccagcgaatgacgcgcgaagaatcttgagatcc 1462  
 |||||  
 Db 241 gaacggcaccacaagaagaagtgccagcgaatgacgcgcgaagaatcttgagatcc 300  
 QY 1463 ttgtgttcaaaagtcctccagcagcagccacacacccgcgaactaaagttcagagagaaga 1522  
 |||||  
 Db 301 ttgtgttcaaaagtcctccagcagcagccacacacccgcgaactaaagttcagagagaaga 360  
 QY 1523 caaccataaagaacacagggatggaagaagaagtaacctttttgaacctggtctggg 1582  
 |||||  
 Db 361 caaccataaagaacacagggatggaagaagaagtaacctttttgaacctggtctggg 420  
 QY 1583 atgaataatggagctagtaataaagaagatagttcagatgaccttatctaatgatacagc 1642  
 |||||  
 Db 421 atgaataatggagctagtaataaagaagatagttcagatgaccttatctaatgatacagc 480  
 QY 1643 agctgcctgtggaatcttccatctgtcccgagagtcgcgcacagcgtgtagagttatgc 1702  
 |||||  
 Db 481 agctgcctgtggaatcttccatctgtcccgagagtcgcgcacagcgtgtagagttatgc 540  
 QY 1703 aaggaatcacacccaagatttaccagggcagctccgaatcctgccaagccaaggttaa 1762  
 |||||  
 Db 541 aaggaatcacacccaagatttaccagggcagctccgaatcctgccaagccaaggttaa 600

```
OY 1763 ctgccatgataccgcagagctgtgtatgaggggcaccctgcacacag-ccgagaccatt 1821
|||||
DB 601 cgcgcctgtatgacccgagagctgtgtatgaggggcaccctgcacacacgagaccatt 660
OY 1822 ttaagaataacatctcttcagggcagcgtaccccatgagctcttcagagacccttgag 1881
|||||
DB 661 ttaagaataacatctcttcagggcagcgtaccccatgagcctcttcagagacccttgag 720
OY 1882 caactgactatcttcagagctcagaggtccaggttcaggttgaatacaagaact 1934
|||||
DB 721 caactgactatcttcagaggtcagaggtcagaggttcaggttgaatacaagaact 773

RESULT 11
X40001/c
ID X40001 standard; DNA: 825 BP.
XX
XX X40001:
XX
XX 02-JUL-1999 (first entry)
XX
DE Prostate cancer.associated gene.
XX
XX Cancer associated antigen; diagnosis; research; treatment; human;
XX breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
XX prostate cancer; ss.
XX
XX Homo sapiens.
XX
XX W09904265-A2.
XX
XX 28-JAN-1999.
XX
XX 15-JUL-1998; 98WO-US14679.
XX
XX 22-JUN-1998; 98US-0102322.
XX
XX 17-JUL-1997; 97US-0896164.
XX
XX 10-OCT-1997; 97US-0061599.
XX
XX 10-OCT-1997; 97US-0061765.
XX
XX 10-OCT-1997; 97US-0948705.
XX
XX 11-OCT-1997; 97GB-0021697.
XX
XX (LUDWIG INST CANCER RES.
XX
XX Chen Y, Gout I, Gure A, O'Hare M, Ohta Y, Old LJ;
XX Pfeunedschuh M, Sahin U, Scanlan MJ, Stockert E;
XX Tureci O;
XX
XX MPI: 1999-132448/11.
XX
XX
XX New isolated cancer associated nucleic acids and polypeptides -
XX isolated using sera from cancer patients, used to develop products
XX for the diagnosis, monitoring or treatment of cancers
XX
XX Claim 67; Page 629; 787pp; English.
XX
XX The invention relates to a method for diagnosing a disorder characterised
XX by expression of a human cancer associated antigen precursor coded for by
XX a nucleic acid molecule (NAM). The method comprises: (a) contacting a
XX biological sample isolated from a subject with an agent that specifically
XX binds to the NAM, an expression product or a fragment of an expression
XX product complexed with an HLA molecule; and (b) determining the
XX interaction between the agent and the NAM or the expression product as a
XX determination of the disorder. The products and methods can be used in
XX the diagnosis, monitoring, research, or treatment of conditions
XX characterised by the expression of various cancer associated antigens.
XX The invention provides nucleic acid sequences and encoded polypeptides
XX which are cancer associated antigen precursors expressed in human breast
XX cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
XX lung cancer.
XX
XX Sequence 825 BP; 289 A; 152 C; 151 G; 228 T; 5 other;
XX
XX
```

```
Query Match 15.7%; Score 551; DB 20; Length 825;
Best Local Similarity 95.9%; Pred. No. 2,1e-131;
Matches 585; Conservative 0; Mismatches 23; Indels 2; Gaps 2:

OY 2889 ttttaattcattgacctaaccagagacagttcagtgatgacaattgtgtccctctgtg 2948
|||||
DB 791 ttttaatttcttcgtgctaaacccagacnacttccagatgacaaattggggccnctgtgt 732
OY 2949 tcaagctgaac-cagtcctgagacttccaacaccttgaataaactcccccagttataa 3007
|||||
DB 731 tcagctgaacacgctcctgagacttccaacaccttgaataaactcccccagttataa 672
OY 3008 attgacaaatttaggaatttaaacctttagatgacatcttggtccatttcaattcalt 3067
|||||
DB 671 attgacaaatttaggaattttaaactttagatgacatcttggtccattttagttmncatt 612
OY 3068 tttaattttgttaalgcgaacagagac-ttaaatgaacttgaactcctgttttaagatta 3126
|||||
DB 611 ttttaatttggtaaatgcaaacagagactttaaataaacttgcctgttttaagattta 552
OY 3127 ttaaaaaaacattgtgtatctatacaatagcctcttgagagacttagcttccactactac 3186
|||||
DB 551 ttaaaaaaacattgtgtatctatacaatagcctcttgagagacttagcttccactactac 492
OY 3187 aggatatgactcctcagtgatgacatataaacctgcagagtgatttccagagtgctgcat 3246
|||||
DB 491 aggatatgactcctcagtgatgacatataaacctgcagagtgatttccagagtgctgcat 432
OY 3247 actgttaattacatctccatllagagctgaaagaatgacctagcttctgtatacagctg 3306
|||||
DB 431 actgttaattacatctccatllagagctgaaagaatgacctagcttctgtatacagctg 372
OY 3307 tgtgcttctgtatgtgtgtactgtacacagaagtgtgtgaactcagagctctgtgtgtg 3366
|||||
DB 371 tgtgcttctgtatgtgtgtactgtacacagaagtgtgtgaactcagagctctgtgtgtg 312
OY 3367 gtcgcatgaaacctgtgtagccctgcaggttaagtagctgttccatcatctgtttagc 3426
|||||
DB 311 gtcgcatgaaacctgtgtagccctgcaggttaagtagctgttccatcatctgtttagc 252
OY 3427 ctggaatttcttcccatggaatgtaagttaaacttaagtgtgtgtcatcaataatgag 3486
|||||
DB 251 ctggaatttcttcccatggaatgtaagttaaacttaagtgtgtgtcatcaataatgag 192
OY 3487 taactactaaa 3496
|||||
DB 191 taactactaaa 182

RESULT 12
X39993/c
ID X39993 standard; DNA: 794 BP.
XX
XX X39993:
XX
XX 02-JUL-1999 (first entry)
XX
XX Prostate cancer associated gene.
XX
XX Cancer associated antigen; diagnosis; research; treatment; human;
XX breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
XX prostate cancer; ss.
XX
XX Homo sapiens.
XX
XX W09904265-A2.
XX
XX 28-JAN-1999.
XX
XX 15-JUL-1998; 98WO-US14679.
XX
XX
```

```
PR 22-JUN-1998; 9805-0102322.
PR 17-JUL-1997; 9705-0896164.
PR 10-OCT-1997; 9705-0061599.
PR 10-OCT-1997; 9705-0061765.
PR 10-OCT-1997; 9705-0948705.
PR 11-OCT-1997; 9705-0021697.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
XX
XX Chen Y, Gout I, Gure A, O'Hare M, Ohata Y, Old LJ;
PI Pfeundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
XX WPI: 1999-132448/11.
XX
XX New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
XX Claim 67: Page 626; 787pp; English.
XX
XX The invention relates to a method for diagnosing a disorder characterised
XX by expression of a human cancer associated antigen precursor coded for by
XX a nucleic acid molecule (NAM). The method comprises: (a) contacting a
XX biological sample isolated from a subject with an agent that specifically
XX binds to the NAM, an expression product or a fragment of an expression
XX product complexed with an HLA molecule; and (b) determining the
XX interaction between the agent and the NAM or the expression product as a
XX determination of the disorder. The products and methods can be used in
XX the diagnosis, monitoring, research, or treatment of conditions
XX characterised by the expression of various cancer associated antigens.
XX The invention provides nucleic acid sequences and encoded polypeptides
XX which are cancer associated antigen precursors expressed in human breast
XX cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
XX lung cancer.
XX
XX Sequence 794 BP; 273 A; 141 C; 152 G; 206 T; 22 other;
SQ

Query Match 14.3%; Score 501.4; DB 20; Length 794;
Best Local Similarity 96.3%; Pred. No. 1e-118;
Matches 517; Conservative 0; Mismatches 18; Indels 2; Gaps 1;

QY 2962 tctcgacttcaaaacctgaatagctcc--acagltgataaatgtgacaattt 3019
DB 718 TCGTGGACCTTCAAAACCCCTGAATAAGTCTCCCCMCAGTTGNATRAATGGCCNATTT 659
QY 3020 aggaatttaaaccttagatgacatcttggttccattttatttcaattttttgtt 3079
DB 658 AGGAATTTTAAACNTAGATGTTTCATTTGGTCCATTTTATTTCAATTTTATTTTGGT 599
QY 3080 aatgcaaacagagctaaatgaacttgatctctgttttaagaatatttaaaaaacattg 3139
DB 598 AATGCAAAACAGGCTTAAAGAACTTGATCTGTGTTTAAAGATTAATTAATAAAACAATTG 539
QY 3140 tgaattataacatgagctcttgagacttagctttcaactaactaagatatgacttc 3199
DB 538 TGNATCTATACATATGGCTCTTGAAGACTTACGTTTACATACACTACAGATATGATCTC 479
QY 3200 catgtatgcataaaacctgcagagtgatcttccagagtgctcogatactgtaattaca 3259
DB 478 CATGTAGTCATATAAACCTGCAGAGGATTTCCANAGGCTCGATACATGTTAATATACA 419
QY 3260 tctccattgagctgtaaaagaatgaacttaagcttctgtataagctgctgttcttgat 3319
DB 418 TCTCCNTTACGGCTGAAAATAATGACCTACGTTCTGTATACANCTGTGTTGCTTTTGAT 359
QY 3320 gtctgttactgacacaagaagctgtgacactgagagctgctgctgtgtgctcgaatgaaa 3379
DB 358 GTTGTGTTACTGNAACACAGAAGTGTGTGCTGAGGCTCTGCGTGTGTGCTGCGTATGAAA 299
QY 3380 acctgttagccctgcgagtgtaagtaactgttcaatcatattgtttaacgttgaaatttct 3439
|||||
```

```
DB 298 GCCCTGGTACCCCTGCAGATTAACTACTGCTTCATTCATTTAGCTGGAAATTTTCT 239
QY 3440 ccccatggaatgttaagaactgaactgatttgcataaataatgtaactaa 3496
DB 238 CCCCATGGAATGTAAAGTAAGTAAGTGTGTGTCATCAATTAATGTAATACTAAA 182

RESULT 13
279981/C
ID 279981 standard; cDNA; 660 BP.
XX
XX 279981;
XX
XX 07-APR-2000 (first entry)
XX
XX Human colon cancer cell line SW480 cDNA clone SEQ ID NO:65.
XX
XX Human: gene expression product; diagnosis; tumour; colon cancer;
KW colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
KW hyperplasia; ds.
XX
XX Homo sapiens.
OS
XX
XX WO964576-A2.
PN
XX
XX 16-DEC-1999.
PD
XX
XX 09-JUN-1999; 99MO-1B01062.
PF
XX
XX 10-JUN-1998; 9805-0088801.
PR
XX
XX (FARB ) BAYER CORP.
PA
XX
XX Endege WO, Steimann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catlino TJ, Dertl A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;
XX
XX WPI: 2000-087220/07.
XX
XX Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer
XX
XX Claim 1; Page 158; 469pp; English.
XX
XX 279917 to 280766 represent double stranded cDNA clones isolated from the
XX human colorectal adenocarcinoma (colon cancer) cell line SW480. The
XX cDNA clones can be used to generate antisense oligonucleotides which
XX can be used for antisense therapy. Methods and products from the present
XX invention can be used for identifying and/or classifying cancerous cells
XX present in a human tumour, particularly in solid tumours, e.g. carcinomas
XX and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used
XX for developing agents for the diagnosis and treatment of disorders
XX involving unwanted cell proliferation, such as neoplasia, dysplasia or
XX hyperplasia.
XX
XX Sequence 660 BP; 133 A; 172 C; 159 G; 174 T; 22 other;
SQ

Query Match 12.4%; Score 433.8; DB 21; Length 660;
Best Local Similarity 92.4%; Pred. No. 1.9e-101;
Matches 495; Conservative 0; Mismatches 36; Indels 5; Gaps 4;

QY 1315 ccagagtaacagctctcacaagagcgagcctccgcgcgaggaagttgtgagcag 1374
DB 531 CCAAGGTACACAGTCTTCANAGAGCNAAGCCTTCCGNGCGNAAGAGATTGTATACAG 472
QY 1375 gtgaagttggaacacacactgcaagaagaagcgacccaagaaggtggtccaaacgc 1434
DB 471 GTNAAGGTTAGAAACCCCATTTGCANNAAGA--CGGGCACACACAGAAGGTGGCCAAAG-GC 415
QY 1435 atgcagccgagacaatgctgagatcctgtgttcaaatcccgagcgagccacc 1494
|||||
```

Db	414	AATGAG-CGAGAACATGTTGAAGATCTTGTTTAAAGTCCCGAGCCGAG-CCAC	357
Qy	1495	aaaccgactcaagtcagagagagaacacccataaagaacccaagggatggaagaaa	1554
Db	356	AAACCGGCACTCAAGTCAGAGGAGAAAGACACCCATTAAGAAACAGGGATGGAAGAAA	297
Qy	1555	qtaacccttttgaacctgtgctctggggatgaaatgggactagtaataagaagatgag	1614
Db	296	GTAACTTTTGTAAACCTGGCTCTGGGGAGAAAAGGAGCACTGTAATTAAGAGATGAG	237
Qy	1615	ctcagatgaccttcttaagtcacagcaagctgcctgcgcggaaatcttcccaatgctcc	1674
Db	236	TTCAAGATGCTTATCTTAAGTCAATCAGACAGCTTCTGCTGGAATTTTCCATGGTGGCC	177
Qy	1675	gaaggctgcgccagagctgtgaggaagtcagagacatcacacaacaaagatttacagagga	1734
Db	176	GAGGTGCGCCGAGGCTGTAGAGGATTAGTCAAAGACATCACACCAAGATTTTACAGGGCA	117
Qy	1735	gtctcgaatctctgcacaaagccacagtgtaactgcatagtatgccgagagtgctgtatggg	1794
Db	116	GCTCCGAACTCTGCGCAAGGCGCACGGTAACCTGCATGATAGCCCGAGAGATTGTGTATGGG	57
Qy	1795	ggagccctcgcccaagcagcgagaccattttaagaataacatctcttcagggccagct	1850
Db	56	GGCACTCTGCGCCACAGCCGAGACCATTTTAAAGATAACATCTTTCAGGCCACGT	1
RESULT	14		
ID	280269		
XX	280269	standard; cDNA; 727 BP.	
AC	280269;		
XX	07-APR-2000	(first entry)	
DT			
XX			
DE		Human colon cancer cell line SW480 cDNA clone SEQ ID NO:353.	
XX			
KW		Human; gene expression product; diagnosis; tumour; colon cancer;	
KW		colorectal adenocarcinoma; cell line SW480; cell proliferation;	
KW		cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;	
KW		hyperplasia; ds.	
OS		Homo sapiens.	
XX			
PN		MO9964576-A2.	
XX			
PD		16-DEC-1999.	
XX			
PF		09-JUN-1999; 99MO-IB01062.	
XX			
PR		10-JUN-1998; 98US-0088801.	
XX			
PA		(FARB ) BAYER CORP.	
XX			
PI		Endege WO, Steilmann KE, Astle JH, Burgess CC, Bushnell SE;	
PI		Carroll E, Catino TJ, Dertl A, Ford DW, Lewis ME, Monahan JE;	
PI		Schlegel R;	
XX			
DR		WPI: 2000-087220/07.	
XX			
PT		Novel nucleic acids, used to develop products for the diagnosis and	
XX		treatment of disorders involving unwanted cell proliferation, and	
PT		particularly cancers, especially colon cancer	
XX			
PS		Claim 15; Page 273; 469pp; English.	
CC			
XX			
CC		279917 to 280766 represent double stranded cDNA clones isolated from the	
CC		human colorectal adenocarcinoma (colon cancer) cell line SW480. The	
CC		cDNA clones can be used to generate antisense oligonucleotides which	
CC		can be used for antisense therapy. Methods and products from the present	
CC		invention can be used for identifying and/or classifying cancerous cells	
CC		present in a human tumour, particularly in solid tumours, e.g. carcinomas	

CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used  
CC for developing agents for the diagnosis and treatment of disorders  
CC involving unwanted cell proliferation, such as neoplasia, dysplasia or  
CC hyperplasia.

Query Match	11.8%;	Score 412.2;	DB 21;	Length 727;
Best Local Similarity	95.5%;	Pred. No. 7e-96;		
Matches 444;	Conservative 0;	Mismatches 17;	Indels 4;	Gaps 2.

Oy	752	ggaattcttccacatttcagttccacgtttacttataaagtgaactttctgaggag	811
Db	1	ggtaatttccacattccagttccacgtttacttataaagtgaactttctgaggag	60
Oy	812	gacagcaatttaatgbcgaagaagaaacaaagacagctctcgaaacaacagatgtctgcga	871
Db	61	gacagcaatttaatgbcgaagaagaaacaaagacagctctcgaaacaacagatgtctgcga	120
Oy	872	aagcgtttgaagatctcgcagaaatgagccctcgcagagaagctctgagatgtaatggaag	931
Db	121	aagcgtttgaagatctcgcagaaatgagccctcgcagagaagctctgagatgtaatggaag	180
Oy	932	aatccgagaagaagaataatctcaataaactctgaaataagtcgaatgttttgagattgaccta	991
Db	181	aatccgagaagaagaataatctcaataaactctgaaataagtcgaatgttttgagattgaccta	240
Oy	992	aacggaactctgcctctgaaattctcgagcttgcccgagagatctgcccacccacatgaaaga	1051
Db	241	aacggaactctgcctctgaaattctcgagcttgcccgagagatctgcccacccacatgaaaga	300
Oy	1052	actttctgacccaaggtttctggttgaggagtttgctgggggaagctgaagggaagaacaga	1111
Db	301	actttctgacccaaggtttctggttgaggagtttgctgggggaagctgaagggaagaacaga	360
Oy	1112	agatttcaagaagaaaatg-ccgcacatagctgttctctgaaagctgaagaagtctgacgcc	1170
Db	361	agatttcaagaagaaaatgacccgcacatagctgttctctgaaagctgaagaagaatctgacgcc	420
Oy	1171	ctgcctgcagcttgaaac---gagtaaaagcctagatcaatcaaaagaa	1212
Db	421	ctgcctgcagatttgacccgaagcttaagccttanaacttcaaatgaa	465

RESULT	ID	217411	standard; CDNA; 769 BP.
XX	AC	217411;	
XX	DT	12-OCT-1999	(first entry)
XX	DE	Human gene expression product	CDNA sequence SEQ ID NO:4884.
XX	KW	Human; gene; gene expression product; diagnosis; therapy; probe;	
XX	KW	detection; mapping; tissue typing; profiling; forensic; cancer;	
XX	XX	genetic analysis; colorectal cancer; breast cancer; lung cancer;	ss
OS	XX	Homo sapiens.	
XX	PN	WO938972-A2.	
XX	PD	05-AUG-1999.	
XX	PF	28-JAN-1999;	99WO-US01619.
XX	PR	03-APR-1998;	98US-0080666.
XX	PR	28-JAN-1998;	98US-0072910.
XX	PR	24-FEB-1998;	98US-0075954.
XX	PR	31-MAR-1998;	98US-0080114.
XX	PR	03-APR-1998;	98US-0080515.



Fri Apr 6 15:25:38 2001

us-09-316-048-5.rng

---

Page 20



```

|||||
153 CGCAATGACGAGCCGAGAAACATGCTGGAGATCTTGGTTCAAACTCCGCA 202
358 narguinprothrlvsprolaleuylserserlglulysrthproilel 375
203 GGGGCGAGCCGCCAACCCGCACTCAAGTCAGAGGAGAGACCCCATTA 252
375 yslasproglaspgllyarglyvalthrphethegluproglysergly 391
|||||
253 AGAAGACAGGGGATGAGAGAAAGTAACCTTTTGTACCTGCTCTGGG 302
392 ASPGLUASGLYTHRSERASNLYSGLUASGLUPHEARGMETPROTYRLE 408
303 GATGAAAATGGGACTAGTATAAAGAGATGACTCAGATGCTTATCT 352
408 userhslsnglnleuprolagllyleuprometvalprogluvala 425
353 AAGTCATACGACACTGCTGCTGGAAATCTTCCCATGCTGCCGAGGTG 402
425 lagnlaalavalglyvalserglnghlshlstrhlyaspphertharg 441
|||||
403 CCAGAGCTGTAGAGATTAGTCAGAGACATCACCAAGATTTTACCAG 452
442 AlaAlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArgL 458
453 GCAGCTCCCAATCTCTGCCAAGGCCACGTAACCTCATGATAGCCGAGA 502
458 uleuleuylrglylthrsrprothralagluthrileuleuylsasn 475
503 GTTGTGTAATGGGGCACTCGCCACAGCCGACCATTTTAAAGAAATA 552
475 snlleserserghlshlvalprohlsglyproleuthrargproserglu 491
553 ACATCTCTTCAGGCCACAGTACCACCATGACCTCTCAGAGACCTCTGAG 602
492 GluLeuAspTyrLeuSerArgValGlnGlyPheGlnValGluTyrLysAs 508
603 CAACCTGACTATCTTCCAGAGTCCAGGGATTCAGATTGAAATACAAAGA 652
508 ppherpolylysasnasn.lysasnghluPhe.ValSer.LeuIleasnCys 524
653 CTTCCTCCCAAAATCACCAAGAACGAATCGGTATCTTATCAATAGCT 702
524 exSerGlnProProLeuIleSerHisGlyIle.GlyLysAspValGluSe 540
703 CCTCTCAGGCACAT.CTGATCCACATGATATTGGCAGGTGGGAATC 751
540 rcysHls...AspMetAlaIleLeuAsnIleLeuLysLeu.LeuSerGlu 555
752 CTGGCAATGATCATGCTGGCTGGACATCTTAAAGATGGCTGGCGGAA 801
556 leuAspGlnGlnSerThr.GluMetProArgThrGlyAsnGlyProMetS 572
802 GTGGACCCCAAAAGTCAAAATAATCCCAAGAAAGGAGAGGCCAGTT 851
572 exVal 573
852 CGGTG 856
seq_name: gb_est77:BB894295
seq_documentation_block:
LOCUS BB894295 752 bp mRNA EST 29-SEP-2000
DEFINITION 601437666F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3922529 5',
mRNA sequence.
ACCESSION BB894295
VERSION BB894295.1 GT:10356518
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 752)

```

```

AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L10M744 row: d column: 18
High quality sequence stop: 654.
FEATURES
location/Qualifiers
1..752
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:3922529"
/clone_id="NIH_MGC_72"
/lisse_type="melanotic melanoma"
/lab_host="FDH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT 223 a 194 c 192 g 143 t
ORIGIN
alignment_scores:
Quality: 1036.50 Length: 249
Ratio: 4.468 Gaps: 4
Percent Similarity: 93.173 Percent Identity: 89.157
alignment_block:
US-09-316-048-6 x BB894295 ..
Align seg 1/1 to: BB894295 from: 1 to: 752
281 TTTGTLGlnGlyIleAsnProIleSerArgLeuAlaGlnIleGlnAla 297
|||||
3 TATGGCCAGGGAGATCAATCGATTACCGCAGCTGGCCAGATCCAGAGGC 52
297 alyslslysgluylsghluproglyluthrleuLeuthrghluargglyleup 314
|||||
53 AAAAAGAGAGAGAGAGAGAGATACAGCTCTCTCAGAGAGAGAGAGCTCC 102
314 roArgArgArgGluPheValMetGlnValLysValGlyAsnHlstrAla 330
|||||
103 CGGCGCGCAGGGAGTTTGTGATGCAGGTGAAGTTGGAACACACACTGCA 152
331 GluGlyThrGlyThrAsnLysValAlaLysArgAsnAlaIleAlaLys 347
|||||
153 GAAGGAAGCGGCGCAACAAGAGAGTGGCCAAACCGCAATGCACCGAGAA 202
347 nmetleugluileleuGlyPheLysValProGlnArgGlnProThrLysP 364
203 CATGCTGGAGATCTTGGTTTCAAACTCCGCGAGCGGCGAGCCCAAAAC 252
364 roAlaLeuLysSerGluLysThrProIleLysLysProGlyAspGly 380
|||||
253 CCGCACTCAAGTCAGAGAGAGAGAGACACCATTAAGAAACACAG. GATGGA 301
381 ArgLysValThrPhePheGluProGlySerGlyAspGluAsnGlyThrSe 397
|||||
302 AGAAAGATAACTTTTGTAAACCTGGCTCTGGGGATGAAATATGGAGTAg 351
397 rasnLysGluAspGluPheArgMetProTyrLeuSerHisGlnGlnLeup 414
|||||
352 TATTAAGAGAGATGACTCAGGATGCTTATCTTAAGTCATCAGACAGCTGC 401

```







```

/def stage="5 months post-conception"
/lab host="DH10B"
/note="Organ: brain; Vector: pBluescript SK (Stratagene);
Site_1: Site_2: XhoI; Double-stranded cDNA was
prepared from human fetal brain tissue. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence:
5'-GAGAGAGAGAAGACCTCAGAGATCCTTAATTAATTAATCCCCCCCCC-3'
and 3' adaptor sequence:
5'-GAGAGAGAGACTCGAGTGTGTTTTTTTTTTT-3'. The library was
size-selected for >0.5 kb inserts and has an average
insert size estimated at 1.2 kb. This library was
constructed using the CAP-trapper method for full-length
enrichment and has not undergone amplification. Library
was constructed by Dr. Claudio Schneider (LNCiB-Area
Science Park, Trieste, Italy). "
BASE COUNT      130 a      152 c      123 g      188 t      1 others
ORIGIN

alignment_scores:
    Quality: 1004.00      Length: 198
    Ratio: 5.122      Gaps: 0
Percent Similarity: 98.990      Percent Identity: 98.485

alignment_block:
US-09-316-048-6 x AM162002/rev ..

Align seg 1/1 to reverse of: AM162002 from: 1 to: 594

69 GluSerIleThrProThrValGluLeuAsnAlaLeuCysMetIysLeuG1 85
|||||
594 GAAAGCATTAACCCCTACCTAGAACATAANATGACCTGGATGAACCTGG 545
|||||
85 YLysLysProMetTyrLysProValAspProTyrSerArgMetGlnSerT 102
|||||
544 AAAAAACCCCAATGTATAACCTGTGACCTTACTCTCGATGACAGTCCA 495
|||||
102 hTyrAsnTyrAsnMetArgIyGlyAlaTyrProProArgTyrPheTyr 118
|||||
494 CCTATTAATACTAACAGATGAGAGAGGCTTATGCCCGAGGTACTTTTAC 445
|||||
119 ProPheProValProProLeuLeuTyrGlnValGluLeuSerValGlyG1 135
|||||
444 CCATTTCACAGTCCACCTTACTTATCAAGTCGAAGCACTTCTGTGGAGG 395
|||||
135 yGlnGlnPheAsnGlyLysGlyLysThrArgGlnAlaAlaLysHisAspA 152
|||||
394 ACAGCAATTTAATGGCAAGAAAGACAAGACGGCTGGAAACACAGATG 345
|||||
152 IaAlaAlaLysAlaLeuArgIleLeuGlnAsnGluProLeuProGluArg 168
|||||
344 CTGCTGCCAAGCGTTGAGAGATCCTGCAAAATGAGCCCTGCCAGAGAG 295
|||||
169 LeuGluValAsnGlyArgGluSerGluGluGluAsnLeuAsnLysSerG1 185
|||||
294 CTGAGAGTAATGGAAGACAATCCGAAGAAAGAAATCTCAATTAATCTGA 245
|||||
185 uIleSerGlnValPheGluIleAlaLeuLysArgAsnLeuProValAsp 202
|||||
244 AATAATCTCAAGTGTTGAGATTTGCACCTTAACCGAAGCACTTGCCTGAAT 195
|||||
202 heGluValAlaArgGluSerGlyProProHisMetLysAsnPheValThr 218
|||||
194 TCGAGGTGGCCGCGAGAGTGGGCCACCCACACTGAAGAAGCTTTGTGACC 145
|||||
219 LysValSerValGlyLubPheValGlyGluGluGlyLysSerLysLys 235
|||||
144 AAGGTTTCGTTGGGAGATTGTGGGGGAAGGGAAGGGAAGAAAGCAAGAA 95
|||||
235 sIleSerLysLysAsnAlaAlaAlaValLeuGluGluLeuLysLysL 252
|||||
94 GATTTCAAAGAAAATGCGCCCATACTCTGTTCTTGAGAGCTGAAGACG 45

```

```

seq_name: gb_est69:BE300155
seq_documentation_block:
LOCUS BE300155 829 bp mRNA EST 20-JUL-2000
DEFINITION 60094460271 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2960593 3',
RNA sequence.
ACCESSION BE300155
VERSION BE3000155.1 GI:9183903
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 829)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMLN at: image.lnl.gov
Plate: LCM54 row: d column: 02
High quality sequence start: 41
High quality sequence stop: 799.
Location/Qualifiers
1..829
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2960593"
/clone_lib="NIH_MGC_17"
/rusue_type="Thadomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: muscle; Vector: pORF7; Site_1: ECORI;
Site_2: XhoI; cDNA made by oligo-dT priming.
directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GCCACGAG(G). Size-selected >500bp
for average insert size 1.kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```

|||||
573 .....GAAGCATACCCCTACTGTAGAACTAAATGACATGTCATGA 529
84 LeuGlyLysProMetLysProValAspProTyrSerArgMetG 100
|||||
528 CTTGGAAAAAACCAATGTAAGCCGTGTGACCTTACTCTCGAGTGA 479
100 nSerThrTyrAsnTyrAsnMetArgGlyAlaTyrProProArgTyrP 117
|||||
478 GTCCACCTATACTACACATGAGAGAGTGTATCCCCGAGGACT 429
117 heTyrProPheProValProProLeuLeuTyrGlnValGluLeuSerVal 133
|||||
428 TTTACCATTTCCAGTTCACCTTACTTATCAAGGCAACTTCTGTG 379
134 GlyGlyGlnGlnPheAsnGlyLysGlyLysThrArgGlnAlaAlaLysH 150
|||||
378 GAGGAGACAGCAATTTAATGGCAAGAACAGACAGAGCTCGAACA 329
150 sAspAlaAlaAlaLysAlaLeuArgGlyLeuGlnAsnGluProLeuProG 167
|||||
328 CGATGCTGCTGCCAAGCCGTTGAGATCCGCAAGATGAGCCCTGCAG 279
167 LuArgLeuGluValAsnGlyArgGluSerGluGluGluAsnLeuLys 183
|||||
278 AGAGCTGAGGTGAATGGAAGAAATCCGAAAGAAATCTCAATAAA 229
184 SerGluLeuSerGlnValPheGluLeuAlaLeuLysArgAsnLeuProVa 200
|||||
228 TCTGAATATGATCAAGCTGTTGAGATTGCACTTAAACGGAACTTGCTGT 179
200 LAsnPheGluValAlaArgGluSerGlyProProHISMetLysAsnPheV 217
|||||
178 GAATTTGAGAGTGGCCGGGAGAGTGGCCACCCACATGAACACTTGTG 129
217 alThrLysValSerValGlyGluPheValGlyGluGlyLysSer 233
|||||
128 TGACCAAGGTTTCTGTTGGGAGCTTCTGGGGAGGTGAAGGAAAAAGC 79
234 LysLysLysLeuSerLysAsnAlaAlaAlaValLeuGluGluLeuLys 250
|||||
78 AACAAATTTCAGAAAGAAATGCCCATGATGCTGTCTGAGAGAGCTGAA 29
250 s 250
28 G 28
seq_name: gb_est43:AW320695
seq_documentation_block:
LOCUS AW320695 633 bp mRNA EST 25-JAN-2000
DEFINITION uc022b10.y1 NCI CGAP Mam6 Mus musculus cDNA clone IMAGE:2609851 5'
similar to TR:095793 095793 STRAUZEN PROTEIN.; mRNA sequence.
ACCESSION AW320695
VERSION AW320695.1 GI:6750239
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 633)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be

```

```

found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
MGI:1023711
Seq primer: -40RP from Gibco
High quality sequence stop: 356.
FEATURES
source
1. 633
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image="IMAGE:2609851"
/clone_lib="NCI-CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT 176 a 167 c 163 g 127 t
ORIGIN
alignment_scores:
Quality: 941.00 Length: 206
Ratio: 4.801 Gaps: 0
Percent Similarity: 95.146 Percent Identity: 86.893
alignment_block:
US-09-316-048-6 x AW320695 ..
Align seg 1/1 to: AW320695 from: 1 to: 633
358 GlnArgGlnProThrLysProAlaLeuLysSerGluGluLysThrProIl 374
|||||
16 CAGCGCGAGCCTGCCAAGCCAGCAGCTCAATTCAGAGAGAACTCCAGT 65
374 eLysLysProGluAspGlyArgLysValThrPhePheGluProGlySerg 391
|||||
66 AAGAAACCCAGAGAGAGAGAAAGTAAACGTTTGTGAACCTAGCCCTG 115
391 LysAspGluAsnGlyThrSerAsnLysGluAspGluPheArgMetProTyr 407
|||||
116 GCGATGAATGCACTAGTAAACAGAGAGAGAGTTCAGAGATGCCCTAT 165
408 LeuSerHisGlnLeuProAlaGlyIleLeuProMetValProGluVal 424
|||||
166 CTTAGGCATCAGCAGCTGCCAGCTGGAAATTCCTCCATGTGTCGGAAGT 215
424 LAlaGlnAlaValGlyValSerGlnGlyHisHISThrLysAspPheThra 441
|||||
216 TGCCAGGCTGTGGGGTTAGTAAAGGACACACACCAAGATTTCACCA 265
441 rGAlaAlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArg 457
|||||
266 GGGCAGCTCCAAATCTCTGCCAAGCAAGGTAACTGCGATGATAGCCGA 315
458 GluLeuLeuTyrGlyLysThrSerProThrAlaGluThrIleLeuLysAs 474
|||||
316 GAGTTGTTGACGGGGGCACTCGCCACAGCCGAGACCATTTTAAAGAG 365
474 nAsnLysSerSerGlyHisValProHisGlyProLeuThrArgProSerg 491
|||||
366 TAACATCTCTTCAGGCGACGTACCCGATGAGCACTCCACAGACCCCTG 415
491 LuGlnLeuAspTyrLeuSerArgValGlnGlyPheGlnValGluTyrLys 507
|||||
416 AGCACTGACTACTCTTTCAGAGCCAGGAGATTCAGGTTGAATCAAA 465
508 AspPheProLysAsnAsnLysAsnGluPheValSerLeuIleAsnCysSe 524
|||||

```



TITLE WashU-NCI human EST Project  
JOURNAL Unpublished (1997)  
COMMENT Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Seq primer: -400P from Glibco  
High quality sequence stop: 455.

## FEATURES

source

```
1..564
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2784225"
/clone_lib="Schneider fetal brain 00004"
/sex="male"
/tissue_type="frontal lobe"
/dev_stage="5 months post-conception"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescript SK (Stratagene);
Site_1: SstI; Site_2: XhoI; Double-stranded cDNA was
prepared from human fetal brain tissue. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence:
5'-GAGAGAGAGAGAGAGCTCAAGATCTTATTAAATTAATCCCCCCCCCCC-3'
and 3' adaptor sequence:
5'-GAGAGAGAGAGAGAGCTCAAGATCTTATTAAATTAATCCCCCCCCCCC-3'.
The library was
size-selected for >0.5 kb inserts and has an average
insert size estimated at 1.2 kb. This library was
constructed using the CAP-trapper method for full-length
enrichment and has not undergone amplification. Library
was constructed by Dr. Claudio Schneider (LNCIB-Area
Science Park, Trieste, Italy)."
```

BASE COUNT 125 a 147 c 117 g 175 t  
ORIGIN

alignment\_scores: Quality: 926.00 Length: 188  
Ratio: 5.060 Gaps: 0  
Percent Similarity: 97.340 Percent Identity: 96.277

alignment\_block:

US-09-316-048-6 x AW163206/rev ..

Align seg 1/1 to reverse of: AW163206 from: 1 to: 564

```
79 AAlaLeuCyMetLysLeuGlyLysPROMeTyrLysProValAsprr 95
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
564 GCACTGTGATACACTTGGAAAAAACCAATGATTATAGCCTGTGACCT 515
95 oTyrSerArgMetGlnSerThrtYrasnTyrAsnMetArgGlyAlaIat 112
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
514 TTACTCTCGATGATCACTCCACCTATTAACCAATGATGAGAGGAGCTT 465
112 yTProBraTgTyrPhetyrProPheProValProProLeuLeuTyrGln 128
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
464 ATCCCCCGAGGTACTTTTACCCATTTCCAGTTCCACCTTACTTTATCAA 415
129 VAIGluLeuSerValGlyGlyGlnGlnPheAsnGlyLysGlyThrAr 145
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
414 GTGGAACTTCTGTGGAGACAGCAATTTAAATGCAAGGAAAGAACAG 365
145 gGluAlaAlaLysHisAspAlaAlaAlaLysAlaLeuArgGileLeuGln 162
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
364 ACAGGCTGGGAAACACGATGCTCTGCCAAAGCGTTGAGGATCTTCACA 315
162 snGluProLeuProGluArgLeuGluValAsnGlyArgGluSerGluGlu 178
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
314 ATGAGCCCTGCCGAGAGAGGCTGGAGGTGAATGGAAGAGAAATCCGAAGA 265
```

```
179 GluAsnLeuAsnLysSerGluIleSerGlnValPheGluIleAlaLeu 195
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
264 GAAATATCTCAATTAATTCGAAATTAAGTCAAGTGTTCAGATTCACCTAA 215
195 sArgAsnLeuProValAsnPheGluValAlaArgGluSerGlyProProH 212
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
214 ACCGAACTTCCTGTGAATTTTCAGAGTGCCCGGAGAGAGTGGCCACCCC 165
212 IsMetLysAsnPheValThrLysValSerValGlyGluPheValGlyGlu 228
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
164 ACATGAAGAACTTTGTGACCAAGTTTCGGTGGGAGTTGTGGGAA 115
229 GLyGluGlyLysSerLysLysIleSerLysLysAsnAlaAlaIleAlaVal 245
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
114 GCGAAGCGGAAAGCAAGAGATTTCAAGAAAAAATTCGCCCATAGCTGT 65
245 lLeuGluGluLeuLysLysLeuProProLeuProAlaValGluArgVal 262
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
64 TCTTGAGAGAGCTGAAGAGTTACCGCCCTGCTGCAGTTGAACGAGTAA 15
262 ySProArgIleLys 266
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
14 AGCCTGAAAAAAA 1
```

seq\_name: gb\_est73:BE617223

seq\_documentation\_block:

LOCUS BE617223 749 bp mRNA EST 24-AUG-2000  
DEFINITION 601441906P1 NIH\_MGC\_65 Homo sapiens cDNA clone IMAGE:3846265 5',  
mRNA sequence.

ACCESSION BE617223

VERSION BE617223.1 GI:9888161

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

<http://image.lnl.gov>

Plate: L10M545 row: k column: 02

High quality sequence stop: 705.

FEATURES

source

```
1..749
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3846265"
/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
```

BASE COUNT 244 a 166 c 223 g 116 t  
ORIGIN

alignment\_scores: Quality: 925.00 Length: 254  
Ratio: 3.903 Gaps: 7

Percent Similarity: 93.307 Percent Identity: 90.551  
 alignment\_block:  
 US-09-316-048-6 x BE617223 ..

Align seg 1/1 to: BE617223 from: 1 to: 749

```

143 LysThrArgGlnAlaAlaLysAspAlaAlaLysAlaLeuArgI 159
|||||
1 AAGACAGACAGCGCTGGAAACAGATGCTGCTCCCAAGCGCTTGAGGA 50
159 LysGlnAsnGluProLeuProGluArgLeuGluValAsnGlyArgGlu 175
|||||
51 TCTGTGAGATGAGCCCTGCCAGAGAGCTGGAGGTGAATGAGAGAA 100
176 SerGluGluLysLeuAsnLysSerGluLysSerGlnValPhe.GluI 192
|||||
101 TCCGAAAGAAATAATCTCAATTAATCTGAATAATGTCACGTGTTGGAGA 150
192 LeuAlaLeuLysArgAsnLeuProValAsnPheGluValAlaArgGluSer 208
|||||
151 TTGCACCTAAACGAACTTGCTCTGTAATTCGAGGTGCGCCGGAGAGT 200
209 GlyProGlnHisMetLysAsnPheValThrLysValSerValGlyGluPh 225
|||||
201 GGCCACCCACATGAGAACTT.GTGACCAAGGTTGCGTTGGGAGTT 248
225 eValGlyGluGlyGluLysSerLysLysLysSerLysAsnAla 242
|||||
249 .GTGGGGGAGGTGAAGGAAAGCAAGAAAGATTTCAAAGAAATCCG 298
242 LaIleAlaValLeuGluLysLysLysLysLeuProProLeuProAlaVal 258
|||||
299 CCATAGCTGTTCTTGAGAGACTGAAGAGTTACCGCCTG.CCTGCAGTT 347
259 GluArgValLysProArgIleLysLysLysThrLysProIleValLysPr 275
|||||
348 GAACGAGTAAAGCTGATCAATCAAAAAGAAACCAACCATGTCAGAGCC 397
275 OGlnThrSerProGluThrGlyGlnGlyIleAsnProIleSerArgLeuA 292
|||||
398 ACAGACAGCGCCAGAAATATGCGCAGGAGT.C.AATCGATTAGCGGACTGG 446
292 LactInIleGlnAlaLysLysGluLysGluProGluThrThrLeuLeu 308
|||||
447 CCCGATCCAGCAGCAAAAAAGAGAGAGCCAGACTACCGCTCCTC 496
309 ThrGluArgLysLeuProArgArgArgGluPheValMetGlnValLysVa 325
|||||
497 ACAGAGCGAGGCTC.CGCGCGCGCAGGAGGTT.GTGATGCAAGTGAAGGT 544
325 LysAsnHisThrAlaGluLysThrGlyThrAsn.LysLysValAlaLys 341
|||||
545 GGGAAACCACTGCAAGAGAGCGGCGACCAAAAGAGTGGCCAAAG 594
342 ArgAsnAlaAlaGluAsnMetLeuGluIleGluGlyPhe..LysValPro 357
|||||
595 CGCAATGACAGCGGAAACATGCTGAGATCTTGTTTCAAAAGTCCCG 644
358 GlnArgLysProThrLysProAlaLeuLysSerGluLysThrProI 374
|||||
645 CAGCGGAGCGCAACAAC...CGGACTCAGTCAGAGAGAGAACACCAT 691
374 eLysLys.ProGlyAspGlyArgLysValThrPhePheGluProGlySer 390
|||||
692 A...AGAAACGGGGATGAGACACGTAACTTTTGGAACTGGCGCG 738
391 Gly 391
|||||
739 GGA 741

```

seq\_name: gb\_est75:BE781223

# seq\_documentation\_block:

LOCUS BE781223 673 bp mRNA EST 20-SEP-2000  
 DEFINITION 601469159f1 NIH\_MGC\_67 Homo sapiens cDNA clone IMAGE:3872247 5',  
 mRNA sequence.  
 ACCESSION BE781223  
 VERSION BE781223.1 GI:10202421  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 (bases 1 to 673)  
 NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL  
 Unpublished (1999)  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert\_Strausberg@nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLC613 row: e column: 16  
 High quality sequence stop: 644.  
 Location/Qualifiers

## FEATURES

### source

1.673  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3872247"  
 /clone\_lib="NIH\_MGC\_67"  
 /tissue\_type="retinoblastoma"  
 /lab\_host="DH10b (phage-resistant)"  
 /note="Organ: eye; Vector: pCMV-Sport6; site\_1: NotI; site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Library constructed by Life Technologies."

BASE COUNT 241 a 156 c 178 g 98 t  
 ORIGIN

## alignment\_scores:

Quality: 908.50 Length: 227  
 Ratio: 4.635 Gaps: 4  
 Percent Similarity: 86.344 Percent Identity: 85.022

## alignment\_block:

US-09-316-048-6 x BE781223 ..

Align seg 1/1 to: BE781223 from: 1 to: 673

```

182 AsnLysSerGluLysSerGlnValPheGluIleAlaLeuLysArgAsnLe 198
|||||
1 AATTAATCTGAAATTAAGTCAAGTGTGAGATTCACCTTAACGAACTT 50
198 uProValAsnPheGluValAlaArgGluSerGlyProProHisMetLysA 215
|||||
51 GCCGTGTAATTTCAGAGTGCGCCGGAGAGTGCGCCACCCACATGAAGA 100
215 snPheValThrLysValSerValGlyGluPheValGlyGluGlyGlu 231
|||||
101 ACTTTGTACCAAGCTTTCGTTGGGAGGTTGTGGGGAAGGTAAGG 150
232 LysSerLysLysLysSerLysAsnAlaAlaIleAlaValLeuGluI 248
|||||
151 AAAAGCAAGAGATTTCAAAGAAAAATCCGCCATAGCTGTTTGAGAGA 200
248 uLeuLysLysLeuProProLeuProAlaValGluArgValLysProArgI 265
|||||
201 GCTGAAGAAGTTACCGCCCTGCTGAGTGAACGAGTAAGCTTAAGAA 250

```

265 lelyslysthrlystProileVallystProglInthSerProglutyr 281  
 |||||||  
 251 TCMAAAGAAACCAAAACCATGTGATGACAGCAGACAGAACCCAGATAT 300  
 |||||||  
 282 GlyGlyGlyIYleasnProIleSerArgleuAlaGlnIleGlnGlnAla 298  
 |||||||  
 301 GGGCAGGGGATCAATCCGATTACCCAGCTGGCCAGATCCAGCAGCAGAA 350  
 |||||||  
 298 slystlystlystProglutyrThrleuLeuThrGluArgGlyLeuPro 315  
 |||||||  
 351 AAGGAGGAAGAGACAGAGTACAGCTCTCCACAGAGCCGAGGCTCCGC 400  
 |||||||  
 315 rgrgrgrglupheValmetGlnVallystValGlyAsnIsthrAlaGlu 331  
 |||||||  
 401 GCGCAGGGGATGTGTGATGACAGTGAAGTTGGAAACCACTGGAGAA 450  
 |||||||  
 332 GlyThrGlyThrAsnlystlystValAlaIyAsnAlaAlaGluAsnMe 348  
 |||||||  
 451 GGAAACGGCCACCAACAAAGAGTGGCCAAAGCCCAATGCAGCCGAAACAT 500  
 |||||||  
 348 tleuglulleugly. PhelystValProglInArgGlnProthlyst..P 364  
 |||||||  
 501 GGTGAGATCTTGGCTTCAAGTCCCGAGGCGCAGCCACCAAAACC 550  
 |||||||  
 364 rolaLeuIySserGluGlyIySthProIlelyst..... 375  
 |||||||  
 551 CAAGACACCCCCCGAAAAAATAAAGAAAGAAACATATAAATAA 600  
 |||||||  
 376 .....LysProGlyAspGlyArgly 382  
 |||||||  
 601 AACCAAAATAAACACCCCAACATTAACAAACAAAGGGGGGGGGCC 650  
 |||||||  
 382 sValThrPhePheGluProGlySergly 391  
 |||||||  
 651 AAAAAAGTTC.....CCCGGGGGGGGC 672

seq\_name: gb\_est3:AA206573

seq\_documentation\_block:

LOCUS AA206573 573 bp mRNA EST 12-MAR-1998  
 DEFINITION zq51f08.t1 StrataGene neuroepithelium (#937231) Homo sapiens cDNA  
 Clone IMAGE:645159 5' similar to SW:STA0\_DROME P25159 MATERNAL  
 EFFECT PROTEIN STAUFEN.; mRNA sequence.

ACCESSION AA206573

VERSION AA206573.1 GI:1801953

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Matra, M., Martin,  
 J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,  
 White, Y., Wylie, T., Waterston, R. and Wilson, R.

WashU-NCI human EST Project

Unpublished (1997)

TITLE

JOURNAL

COMMENT

Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: est@wustl.wustl.edu

This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (InfoImage.lnl.gov) for further information.

Insert Length: 3237 Std Error: 0.00

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 325.

Location/Qualifiers

FEATURES

1..573  
 /organism="Homo sapiens"  
 /db\_xref="GDB:5215410"  
 /db\_xref="taxon:9606"

/clone="IMAGE:645159"  
 /clone\_lib="Stratagene neuroepithelium (#937231)"  
 /dev\_stage="Ntera-2/RA neuroepithelial cells"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /note="vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 XhoI; Cloned unidirectionally. Primer: Oligo dT, NT2  
 cells (Ntera-2/cl.D1) induced with Retinoic Acid for 24  
 hours. Average insert size: 1.5 kb; Uni-ZAP XR Vector; -5'  
 adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor  
 sequence: 5' CTCGAGTTTCTTTTCTTTT 3'"

BASE COUNT 178 a 118 c 145 g 129 t 3 others  
 ORIGIN

alignment\_scores:  
 Quality: 882.00 Length: 191  
 Ratio: 4.642 Gaps: 0  
 Percent Similarity: 99.476 Percent Identity: 96.859

alignment\_block:

US-09-316-048-6 x AA206573 ..

Align seg 1/1 to: AA206573 from: 1 to: 573

69 GluserIleThrProthryValGluLeuAsnAlaLeuCyMetlystleugl 85  
 |||||||  
 4 GAAGCATTAACCCCTTACTGATGACTAAATGCATCTGATGATAACTTGG 53  
 |||||||  
 85 ylystlystPrometlystlystProValaspprotyrSeraArgMetGlnSert 102  
 |||||||  
 54 AAAAAAACCAATGTATTAACCTGTGACCTTACTCTCGGATGCACTGCA 103  
 |||||||  
 102 htyrtnslyrAsnMetArgGlyAlaIleTyProProArgyrrPheTy 118  
 |||||||  
 104 CCTATACTACACATGAGAGAGGCTTATCCCGAGGATCTTTTAC 153  
 |||||||  
 119 ProPheProValProProleuLeuTyrglnValGluLeuSerValGly 135  
 |||||||  
 154 CCATTTCCAGTTCCACCTTACTTATCAAGTGAACCTTCTGTGGAGG 203  
 |||||||  
 135 yglnglnPheasnGlyLyGlyLystThraArgGlnAlaIleLySAsp 152  
 |||||||  
 204 ACAGCAATTAATGCGCAAGAAAGCAAGACAGCTCGCAAAACAGATG 253  
 |||||||  
 152 lalaIalaIyAlaLeuArgIleLeuGlnAsnGluProleuProgluArg 168  
 |||||||  
 254 CTGCTGCCAAAGCGTTGAGGATCTCGAGATGAGCCCTGCCAGAGAG 303  
 |||||||  
 169 LeuGluValAsnGlyArgGluSerGluGluGluAsnLeuAsnlySergl 185  
 |||||||  
 304 CTGAGAGTGAATGAGAGACATCCGAAGAAATAATCTCATTAATCTGA 353  
 |||||||  
 185 uIleSerglnValPheGluIleAlaLeuIySArgAsnLeuProValAsp 202  
 |||||||  
 354 AATTAATCAAGTGTGAGATTGCACTTAACGGAACCTTGCCGTGTAAT 403  
 |||||||  
 202 heGluValAlaArgGluSerGlyProProhIsmetyAsnPheValThr 218  
 |||||||  
 404 TCGAGGTGGCCCG. GAGAGTGGCCACCCACATGAGAACTTTGAGAC 452  
 |||||||  
 219 LySValSerValGlyLupheValGlyGlyGlyGlyLySserlystly 235  
 |||||||  
 453 AAGGTTTGGGTTGGGAG. TTTGTGGGGGAAGGGAAGGAAAGCCAGAA 501  
 |||||||  
 235 sIleSerLySAsnAlaIleAlaIleAlaValleuglGluLeuLySly 252  
 |||||||  
 502 GATTTTC. AAGAAAAATGCCCATACCTGTTGAGAGCTGAAGAACT 550  
 |||||||  
 252 euProProleuProAlaValGlu 259  
 |||||||  
 551 TTCNGCCCTCGCTGCATTAATGAA 573

seq\_name: gb\_est77:BB872137



```

seq_documentation_block:
LOCUS      BE872137      852 bp      mRNA      EST      27-SEP-2000
DEFINITION 601446238P1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3850496 5',
            mRNA sequence.
ACCESSION  BE872137
VERSION    BE872137.1 GI:10320913
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE  1 (bases 1 to 852)
            NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
JOURNAL    Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LICM556 row: k column: 09
            High quality sequence stop: 637.
FEATURES
    source
        1..852
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:3850496"
            /clone_lib="NIH_MGC_65"
            /tissue_type="adenocarcinoma"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: Colon; Vector: pCMV-SPORT6; Site: 1: NotI;
            Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
            Average insert size 1.8 kb. Library constructed by Life
            Technologies."
BASE COUNT  275 a      215 c      218 g      144 t
ORIGIN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852

```

```

seq_name: gb_est72:BE541462
seq_documentation_block:
LOCUS      BE541462      673 bp      mRNA      EST      09-AUG-2000
DEFINITION 601067913P1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:3454179 5',
            mRNA sequence.
ACCESSION  BE541462
VERSION    BE541462.1 GI:9770107
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE  1 (bases 1 to 673)
            NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
JOURNAL    Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: Incyte Genomics, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LHM8438 row: j column: 04
            High quality sequence stop: 637.
FEATURES
    source
        1..673
            /organism="Homo sapiens"

```



```
288 ILeSerArgLeuAlaGlnIleGlnGlnAla.LysLysGluLysGluProG 304
|||||
153 ATTAGCCGACTGCGCCAGATCCAGCAGGCAAAAAAGAGAGAGAGCCAC 202
|||||
304 LuTyTrhLeuLeuThrGluArgGlyLeuProArgArgArgGluPheVal 320
|||||
203 AGTACAGCCTCTCACAGCGAGGCGCTCCCGCGCGAGGAGTTGTGA 252
|||||
321 .MetGlnValLysValGlyAsnHisThrAlaGluGlyThrGlyThrAsnL 337
|||||
253 GATGACAGGTAGAGTTGGAAMCCACACTGCGAGAAGAAAGGCGCACAA 302
|||||
337 yslLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGluIleLeuGly 353
|||||
303 AGAAGGTGGCCCAAGCGCAATGCACGCCAGACATGCTGGAGATCCTTGGT 352
|||||
354 PheLysValProGlnArgGlnProThrlsProAlaLeuLysSerGluG 370
|||||
353 TTCAAAGTCCCGCAGGCGCAGCCACCACCCGCTCAAGTCAGAGGA 402
|||||
370 ulysThrProIleLysLysProGlyAspGlyArgLysValThrPhePheG 387
|||||
403 GAAGACACCCCTTAAGAAACCAGGGGATGGAGAAGAAAGTAACCTTTT 452
|||||
387 LuProGlySerGlyAspGluAsnGlyThrSerAsnLysGluAspGluPhe 403
|||||
453 AACCTGGCTGGGGATGAACTAGGACTAGTAAAGAGAGATGAGTTTC 502
|||||
404 ArgMetProTyTrhLeuSerHisGlnGlnLeuProAlaGlyIleLeuPro 420
|||||
503 AGGATGGCTTATCTAGTCATCAGCAGCTGCTGCTGGAATCTTCCCAT 552
|||||
420 tValProGluValAlaGlnAlaValGlyValSerGlnGlyHisHisThrL 437
|||||
553 GGTACCGACGACTCCG.CAGGCTGAGAGGTAGTCAGAGACATCACACCA 601
|||||
437 ysAspPheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAla 453
|||||
602 CAGATTGA.ACAGGCCAGCATCCGAATCCTGCCAAGACACGACTACTGCA 650
|||||
454 MetIleAlaArgGluLeuLeuTyTrhGlyGlyThrSerProThrlaGluTh 470
|||||
651 CAAAGATAGCCCGAAGTGGGAGAGGGGCGACCTGACCA..... 689
|||||
470 rIleLeuLysAsnAsnIleSerSerGlyHisValProHisGly 484
|||||
690 .....CAGCGAAACATATCATCAGACACACCTCTCTCAGGGGC 725
|||||
```





1039 ...GGCTGCTGCACCGT 1025

seq\_name: /cgn2\_6/ptodata/2/lna/5B\_COMB.seq:US-08-718-661-1

seq\_documentation\_block:

Sequence 1, Application US/08718661

Patent No. 5876972

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: Nucleic acid molecules coding for mammalian

TITLE OF INVENTION: tumor suppressor proteins and methods for their isolation

NUMBER OF SEQUENCES: 15

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)

APPLICATION NUMBER: US/08/718, 661

CLASSIFICATION: 530

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2790 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA to mRNA

FEATURE:

NAME/KEY: CDS

LOCATION: 542..2545

US-08-718-661-1

alignment\_scores:

Quality: 101.50 Length: 47

Ratio: 2.985 Gaps: 4

Percent Similarity: 72.340 Percent Identity: 42.553

alignment\_block:

US-09-316-048-27 x US-08-718-661-1/rev ..

Align seg 1/1 to reverse of: US-08-718-661-1 from: 1 to: 2790

1 ThrAlaCysAlaAlaLthrCysThrAlaGlyAlaGlyCysGlyCys.. 16

2286 TCAGGCTGCTGCTGAGAGTTGCGCAGCAGCTGTGAGTGCAT 2237

17 .....Cys.....GlyCysGlyCysThrCysAlaGlyAlaG 27

2236 CTGGGCGCTGATCTGCACACTGGAACCTGCTGCATCTGT...GGCTGTG 2190

27 lylGlyGlyThrCysThrAlaGlyThrGlyCysGlyAlaGly 40

2189 GTAGTGGC...TGTGGCACTGGCAGTGTGCTGCGCAGTGGC 2152

seq\_name: /cgn2\_6/ptodata/2/lna/5B\_COMB.seq:US-08-991-300-1

seq\_documentation\_block:

Sequence 1, Application US/08991300

Patent No. 5973225

GENERAL INFORMATION:

APPLICANT: D'OVIDIO, RENATO

APPLICANT: PORCEDDU, ENRICO

APPLICANT: MERCHITELLI, CINZIA

APPLICANT: CARRELLI, LUISA ERCOLI

TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF A GENE

TITLE OF INVENTION: ENCODING A LOW MOLECULAR WEIGHT GLUTENIN

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBION, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

ADDRESS: P.C.

STREET: 1755 S. JEFFERSON DAVIS HIGHWAY

CITY: ARLINGTON

STATE: VA

COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/991,300

FILING DATE: 16-DEC-1997

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: IT MI 96/A 002663

FILING DATE: 19-DEC-1996

ATTORNEY/AGENT INFORMATION:

NAME: OBION, NORMAN F.

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 2264-0201-0X

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-413-3000

TELEFAX: 703-413-2220

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1107 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-991-300-1

alignment\_scores:

Quality: 95.50 Length: 48

Ratio: 3.673 Gaps: 3

Percent Similarity: 54.167 Percent Identity: 41.667

alignment\_block:

US-09-316-048-27 x US-08-991-300-1/rev ..

Align seg 1/1 to reverse of: US-08-991-300-1 from: 1 to: 1107

1 ThrAlaCysAlaAlaLthr.....CysThrAlaGlyAlaGlyCys 13

572 ACTGCGTGTGTTGTTGTCGCAAAATGCTGTTGTCGCGTGAATTGTTG 523

13 s.....GlyGlyCysCysGly...CysGlyCysThrc 23

522 TGTGTTGTTGCGAAATGTTGTTGTTGCGTACTGCTGTTGTTGTT 473

23 ysaAlaGlyAlaGlyGlyGlyThrCysThrAlaGlyThrGlyCys 37

472 GCGCGCAAAATGCTGTAAGTCTGCAAAATGCTGCTGCTGT 429

seq\_name: /cgn2\_6/ptodata/2/lna/6\_COMB.seq:US-08-700-651-1

seq\_documentation\_block:

Sequence 1, Application US/08700651B

Patent No. 6015882

GENERAL INFORMATION:

APPLICANT: PETERSEN, CAROLYN

APPLICANT: LEECH, JAMES

APPLICANT: NELSON, RICHARD, C.

APPLICANT: GUT, JIRI

TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS

TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cytosporidium parvum

FILE REFERENCE: 480.19-4(HV)

CURRENT APPLICATION NUMBER: US/08/700,651B

CURRENT FILING DATE: 1997-08-14

EARLIER APPLICATION NUMBER: 08/415,751

```

: EARLIER FILING DATE: 1995-04-03
: NUMBER OF SEQ ID NOS: 15
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 5163
: TYPE: DNA
: ORGANISM: Cryptosporidium parvum
US-08-700-651-1

alignment_scores:
      Quality: 94.50      Length: 31
      Ratio: 4.109      Gaps: 1
Percent Similarity: 74.194      Percent Identity: 45.161

alignment_block:
US-09-316-048-27 x US-08-700-651-1/rev ..

Align seg 1/1 to reverse of: US-08-700-651-1 from: 1 to: 5163

      7 CysThAlaGlyAlaGlyCysGlyCysCysGlyCysGlyCysThCys 23
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
      865 TGTAGCTGTGCTGTAGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 816

      23 sAlaGlyAlaGlyGlyGlyThCysThAlaGlyThCys 37
      |||:|||||:|||||:|||||:|||||:|||||:|||||
      815 TTGTGTGT.....TGTAGTTGTGTTGTAGTTGT 788

seq_name: /cgn2_6/plodata/2/lna/5_COMB.seq:US-08-928-361B-4

seq_documentation_block:
: Sequence 4, Application US/08928361B
: Patent No. 6071518
: GENERAL INFORMATION:
: APPLICANT: Petersen, Carolyn
: TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
: TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
: TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
: NUMBER OF SEQUENCES: 30
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: PETERS, VERNY, JONES & BIKSA
: STREET: 385 Sherman Avenue, Suite 6
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94306-1840
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/928,361B
: FILING DATE: 12-SEP-1997
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/026,062
: FILING DATE: 13-SEP-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: VERNY, Hana
: REGISTRATION NUMBER: 30,518
: REFERENCE/DOCKET NUMBER: 480,76-1(HV)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-324-1677
: TELEFAX: 650-324-1678
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5163 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)

```

```

US-08-928-361B-4
alignment_scores:
  Quality: 94.50      Length: 31
  Ratio: 4.109      Gaps: 1
  Percent Similarity: 74.194      Percent Identity: 45.161

Alignment block:
US-09-316-048-27 x US-08-928-361B-4/rev ..

Align seg 1/1 to reverse of: US-08-928-361B-4 from: 1 to: 5163

      7  CysThAlaGlyAlaGlyCysGlyGlyCysGlyCysGlyCysThrcy 23
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      864 TGTAGTTGGTGTGAGTTGTTGGTGTGTTGTTGTTGTTGTTGAGTTG 815

      23 salagiAlaGlyGlyGlyThrcCysThAlaGlyThrcGlyCys 37
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
      814 TTGTGCT.....TGTAGTTGGTGTGAGTTGT 787

seq_name: /cgn2_6/plodata/2/ina/6_COMB.seq:US-08-700-651-2
seq_documentation_block:
; Sequence 2, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUY, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; FILE REFERENCE: 480.19-4(HV)
; CURRENT APPLICATION NUMBER: US/08/700,651B
; EARLIER FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 5318
; TYPE: DNA
; ORGANISM: Cryptosporidium parvum
US-08-700-651-2

alignment_scores:
  Quality: 94.50      Length: 31
  Ratio: 4.109      Gaps: 1
  Percent Similarity: 74.194      Percent Identity: 45.161

Alignment block:
US-09-316-048-27 x US-08-700-651-2/rev ..

Align seg 1/1 to reverse of: US-08-700-651-2 from: 1 to: 5318

      7  CysThAlaGlyAlaGlyCysGlyGlyCysGlyCysGlyCysThrcy 23
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      865 TGTAGTTGGTGTGAGTTGTTGGTGTGTTGTTGTTGTTGTTGAGTTG 816

      23 salagiAlaGlyGlyGlyThrcCysThAlaGlyThrcGlyCys 37
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
      815 TTGTGCT.....TGTAGTTGGTGTGAGTTGT 788

seq_name: /cgn2_6/plodata/2/ina/6_COMB.seq:US-08-928-361B-3
seq_documentation_block:
; Sequence 3, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
;

```

```

ADDRESS: Naval Medical Res. & Dev. Cnd.
STREET: Bldg 11, T-12 8901 Wisconsin Ave.
CITY: Bethesda
STATE: Maryland
COUNTRY: USA
ZIP: 20889-5606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/155,888
FILING DATE:

CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spevack, A. David
REGISTRATION NUMBER: 24,743
REFERENCE/DOCKET NUMBER: N.C. 75,851
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 295-6759
TELEFAX: (202) 295-1022
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5552 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
HYPOTHETICAL: NO
ANTI-SENSE: NO

FEATURE:
NAME/KEY: promoter
LOCATION: 1..755
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "promoter"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /label= CMV-IE
OTHER INFORMATION: /note= "this feature acts as a promoter for any
OTHER INFORMATION: downstream DNA sequence."
OTHER INFORMATION: /citation= (12)]

FEATURE:
NAME/KEY: CDS
LOCATION: 933..2367
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 933
OTHER INFORMATION: /function= "protein protective against malaria"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /number= 1
OTHER INFORMATION: /label= II2-GSP
OTHER INFORMATION: /citation= (11)]

PUBLICATION INFORMATION:
AUTHORS: Sedegah, Martha
AUTHORS: Hedstrom, Richard C.
AUTHORS: Hoffman, Stephen L.
TITLE: Vaccination with Plasmodium yoelii CS protein
TITLE: plasmid DNA protects against malaria
JOURNAL: Science
PUBLICATION INFORMATION:
AUTHORS: Cullen, Bryan R.
TITLE: TRANS-ACTIVATION OF HUMAN IMMUNODEFICIENCY
TITLE: VIRUS OCCURS VIA A BIMODAL MECHANISM
JOURNAL: CELL
VOLUME: 46
PAGES: 973-982
DATE: 26 SEP-1986
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 4732
US-08-155-888-1

```

















```

IMAGE:1230042 5', mRNA sequence.
ACCESSION   AA798861
VERSION     AA798861.1
KEYWORDS    GI:2861816
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE   1 (bases 1 to 355)
AUTHORS     Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucada,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Stepoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
TITLE       The WashU-HHMI Mouse EST Project
JOURNAL     Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel.: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:655634
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 249.
FEATURES             source          location/Qualifiers
                     1..355
                     /organism="Mus musculus"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1230042"
                     /clone_id="Stratagene mouse skin (#937313)"
                     /sex="females"
                     /tissue_type="whole skin"
                     /dev_stage="11 weeks old"
                     /lab_host="SDR (kanamycin resistant)"
                     /note="Organ: skin; Vector: pluescript SK-; site_1: EcoRI
                        ; site_2: XhoI; Cloned unidirectionally. Primer: Oligo
                        dT. Whole skin from 11 week old C57BL/6 female mice.
                        Average insert size: 1.0 kb. Uni-ZAP XR Vector. ~5'
                        adaptor sequence: 5' GAAATTCGCACGAG 3'-3' adaptor
                        sequence: 5' CTCGAGATTATTTTATTTTTTTT 3'."
BASE COUNT        19 a              74 c              161 g              101 t
ORIGIN
alignment_scores:
Quality:         117.00           Length:         39
Ratio:           4.034            Gaps:           3
Percent Similarity: 74.359        Percent Identity: 56.410
Alignment_block:
US-09-316-048-27 x AA798861 ..
Align seg 1/1 to: AA798861 from: 1 to: 355
3 CysAlaIaThrcCysAlaGlYrAlaGlYcysGlygLyCysCysGlyCY 19
|||||.....||| |||||.....||||| |||||.....|||||
211 TCGCTGGCGTGCTGT.....GGTTGGTGTCTGCTGGCTGTGGCTG 254
19 s...glYcysThrCysAlaGlYrAlaGlYgLyThrCysThralaGlYT 35
| | | | | | | | | | | | | | | | | | | | | | | | | | |
255 TGGTGGCTGTGGCTGCGGTGGCTGTGGCTGTGGCTGTGGCTGTGGCC. 303
35 hrcLYCysGLyAlaGLy 40
| | | | | | | | | | | | | | | | | | | | | | | | | | |
304 ..GGTTGTGGCTGGCGT 318
seq_name: gb_est21:A1509887
```



```

seq_documentation_block:
LOCUS      293 bp      mRNA      EST      15-MAR-2000
DEFINITION ms32d09.y1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
IMAGE:608657 5', mRNA sequence.
ACCESSION  A1509887
VERSION     A1509887
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 293)
AUTHORS     Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schuck,R., Rittler
            E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
            The WashU-NCI Mouse EST Project 1999
TITLE        Unpublished (1999)
JOURNAL      Contact: Marra M/WashU-NCI Mouse EST Project 1999
COMMENT      Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            This read is a RESQUENCE of a previously sequenced mouse clone
            This read has been verified (found to hit its original self in the
            correct orientation)
            MGI:374089
            Seq primer: -40RP from Gibco
            High quality sequence stop: 272
            POLYA-No.

FEATURES             Location/Qualifiers
     source           1..293
                     /organism="Mus musculus"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:608657"
                     /clone_lib="Stratagene mouse skin (#937313)"
                     /sex="females"
                     /tissue_type="whole skin"
                     /dev_stage="11 weeks old"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1: EcoRI
                     dt. Whole skin from 11 week old C57BL/6 female mice.
                     Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
                     adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor
                     sequence: 5' CTCGAGCTTTT TTTT TTTT TTTT 3'"

BASE COUNT          31 a      69 c      131 g      62 t

ORIGIN
1
alignment_scores:
Quality: 116.50      Length: 45
Ratio: 3.758      Gaps: 2
Percent Similarity: 68.889      Percent Identity: 44.444

alignment_block:
US-09-316-048-27 x A1509887 ..
Align seg 1/1 to: A1509887 from: 1 to: 293

1  ThrAlAcysAlaAaIarhrCysThrAlaGlyAlaGlyCysGlyCysCy 17
   :::::::::::::: :::::::::::::: :::::::::::::: ::
150  TCTAGCTGCGAGAGAGGCTGC...TCGCGGCTGCTGCGGCGGATGCGG 196
   ||||||||| :::::::::::::: :::::::::::::: ::
17  sglyCysGlyCysThrCysAlaGlyAlaGly..... 27
   ||||||||| :::::::::::::: :::::::::::::: ::
197  AGGCTGCGGCTAGCTGCTGCGGATGCTGCGGAGAGGCTGCGGG 246

```

```

28 .....GlyGlyThrCysThrAlaGlyThrGlyCys 37
   ||||| :::::::::::::: :::::::::::::: ::
247  GCTGTGAGAGCGGCTGCTGCGGCTGCGGATGCG 281
   ||||| :::::::::::::: :::::::::::::: ::
seq_name: gb_est11:AA727427

seq_documentation_block:
LOCUS      242 bp      mRNA      EST      02-JAN-1998
DEFINITION vu96a04.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
IMAGE:1209966 5', mRNA sequence.
ACCESSION  AA727427
VERSION     AA727427
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 242)
AUTHORS     Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
            The WashU-HHMT Mouse EST Project
TITLE        Unpublished (1996)
JOURNAL      Contact: Marra M/Mouse EST Project
COMMENT      WashU-HHMT Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:646310
            Seq primer: -28m13 rev1 ET from Amerham
            High quality sequence stop: 200.
            Location/Qualifiers

FEATURES             Location/Qualifiers
     source           1..242
                     /organism="Mus musculus"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1209966"
                     /clone_lib="Stratagene mouse skin (#937313)"
                     /sex="females"
                     /tissue_type="whole skin"
                     /dev_stage="11 weeks old"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1: EcoRI
                     dt. Whole skin from 11 week old C57BL/6 female mice.
                     Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
                     adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor
                     sequence: 5' CTCGAGCTTTT TTTT TTTT TTTT 3'"

BASE COUNT          11 a      43 c      112 g      76 t

ORIGIN
1
alignment_scores:
Quality: 113.50      Length: 42
Ratio: 3.783      Gaps: 2
Percent Similarity: 71.429      Percent Identity: 50.000

alignment_block:
US-09-316-048-27 x AA727427 ..
Align seg 1/1 to: AA727427 from: 1 to: 242

3  CysAlaAaIarhrCysThrAlaGlyAlaGlyCysGlyCys..... 16
   ||:::::::::::: :::::::::::::: :::::::::::::: ::
99  TGTGTGCTGTGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 148
   ||:::::::::::: :::::::::::::: :::::::::::::: ::
17  ....CysGlyCysGlyCysThrCysAlaGlyAlaGlyCysGlyCys 32

```





```

/notice="Organ: skin; Vector: pbluescript SK-; Site: 1; EcoRI
; Site: 2; XhoI; Cloned unidirectionally. Primer: Oligo
dT. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb; Uni-GAP XR Vector: -5'
adaptor sequence: 5' GAATTCGGCAG 3' -3' adaptor
sequence: 5' CTCGAGTCTTTTTTTTTTTTTT 3'"
BASE COUNT      20 a      48 c      85 g      68 t
ORIGIN
alignment_scores:
    Quality: 111.00      Length: 42
    Ratio: 3.700      Gaps: 4
Percent Similarity: 71.429      Percent Identity: 54.762
alignment_block:
US-09-316-048-27 x AA733514 ..
Align seg 1/1 to: AA733514 from: 1 to: 221
1 ThrAlaCysAlaAlaThrcysThrAlaGlyAlaGlyCysGlyCyscy 17
:::||||| :::||||| |||||||
87 ACCAGCTGC...ACCACCTGCAGATCTACCCGCTGCTGTGCTGCTG 133
17 scgLy...CysGlyCysThrCysAlaGlyAlaGlyLy...GlyThrCysT 32
||||| ||||||| |||:::|||||:::||||| ||| |||
134 TGCCTCTCTGTGCTGTGCTGTGCTGTGCGGTGCTGTGCTGTGCTGTG 183
32 hrAlaGlyThrGlyCysGlyAlaGly 40
:::||||| |||||||:::|||||
184 GCTGTGCT...GGCTGTGCTGTGCTG 206
seq_name: gb_est8:AA530753
seq_documentation_block:
LOCUS      AA530753      521 bp      mRNA      EST      22-JUL-1997
DEFINITION vj44ae07.r1 Stratiagene mouse skin (#937313) Mus musculus cDNA clone
IMAGE:931908 5', mRNA sequence.
ACCESSION  AA530753
VERSION    AA530753
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM  Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eulharia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 521)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Teisling,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watsn.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:536828
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 298.
Location/Qualifiers
1..521
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:931908"
/location="Stratiagene mouse skin (#937313)"
/sex="Females"
FEATURES
SOURCE

```

```

/tissue-type="whole skin"
/dev_stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/node="Organ: skin; Vector: pBluescript SK-; Site_1: Eco
/site_2: XhoI; Cloned unidirectionally. Primer: Oligo
dt. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb; uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor
sequence: 5' CTCGACTTTTCTTTTCTTTTCTTTT 3'"
BASE COUNT      62 a      122 c      188 g      149 t
ORIGIN

alignment_scores:
    Quality:      111.00      Length:      42
    Ratio:        3.700      Gaps:      4
Percent Similarity: 71.429      Percent Identity: 54.762

alignment block:
US-09-316-048-27 x AA530753 ..

Align seg 1/1 to: AA530753 from: 1 to: 521

1 ThrAlaCysAlaAlaFhcCysThrAlaGlyAlaGlyCysGlyCysC 17
   :::::||||| :::::|||||
4 AGCAGCTGC..ACCACCTGCAGATACACCGCGCTGTGGTGGTGGT 50
   17 sGly...CysGlyCysThrCysAlaGlyAlaGlyGly...GlyThrCysT 32
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 TGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 100
   ::::: ||| ||||| ||||| |||||
101 GCTGTGCT...GGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 123

seq_name: gb_est29:AU061961

seq_documentation_block:
LOCUS      AU061961          319 bp      mRNA      EST      20-MAY-1999
DEFINITION AU061961 Dictyostelium discoidium SL (H. Urushihara) Dictyostelium
discoidium cDNA clone SLG781, mRNA sequence.
ACCESSION  AU061961
VERSION     AU061961.1  GI:4883065
KEYWORDS
SOURCE      EST.
ORGANISM    Dictyostelium discoidium.
            Dictyostelium discoidium.
            Eukaryota; Dictyostelida; Dictyostelium.
REFERENCE   1 (bases 1 to 319)
AUTHORS    Yoshino, R., Morio, T. and Tanaka, Y.
TITLE      Developmental cDNA in Dictyostelium discoidium
JOURNAL    Unpublished (1997)
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
            Email: d402huesakura.cc.tsukuba.ac.jp
PROJECT     = Dictyostelium discoidium cDNA project in Japan.
            location/Qualifiers
            1. 319
               /organism="Dictyostelium discoidium"
               /strain="AX4"
               /db_xref="taxon:44689"
               /clone="SLG781"
               /clone_lib="Dictyostelium discoidium SL (H. Urushihara)"
               /dev_stage="slug"
BASE COUNT      188 a      102 c      6 g      23 t
ORIGIN

alignment_scores:
    Quality:      110.50      Length:      40
    Ratio:        3.810      Gaps:      1
Percent Similarity: 72.500      Percent Identity: 47.500

```

## alignment\_block:

US-09-316-048-27 x AU061961/rev ..

Align seg 1/1 to reverse of: AU061961 from: 1 to: 319

```

1 ThrAlaCysAlaAlaThrCysThrAlaGlyAlaGlyCysGlyCysC 17
  :::::::::::::: :::: ::::: ::::: :::::
211 TCTGTGTTGTTGTCGGTGTG...TGTGTTGTTGTTGTTGTTGTTG 165
17 sGlyCysGlyCysThrCysAlaGlyAlaGlyClyGlyThrCysThrAlaG 34
  ||||| ||||| ||||| :::: :::::
164 TGGTTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 115
34 1yThrGlyCysGlyAlaGly 40
  || ||||| ||||| :::::
114 GTTGTGCTGTGCTGTGCTGTGCT 95
seq_name: gb_est22:AI614298

```

## seq\_documentation\_block:

```

LOCUS AI614298 381 bp mRNA EST 21-APR-1999
DEFINITION vi83g12.y1 StrataGene mouse skin (#937313) Mus musculus cDNA clone
ACCESSION AI614298
VERSION AI614298.1 GI:4623465
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 381)

```

## REFERENCE

AUTHORS Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schmitt, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

TITLE The WashU-NCI Mouse EST Project 1999

JOURNAL Unpublished (1999)

COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:531102

This read is a RESEQUENCE of a previously sequenced mouse clone. This read has been verified (found to hit its original self in the correct orientation)

Seq primer: -40RP from Gthco  
High quality sequence stop: 380.

## FEATURES

## source

```

1. 381
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:918886"
/clone_lib="StrataGene mouse skin (#937313)"
/sex="females"
/tissue_type="whole skin"
/dev_stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skin; Vector: pBluescript SK-; Site: 1: EcoRI; Site: 2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Whole skin from 11 week old C57BL/6 female mice. Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5' CTCGAGTTT TTT TTT TTT TTT TTT 3'"
BASE COUNT 31 a 98 c 151 g 99 t 2 others
ORIGIN

```

## alignment\_scores:

Quality:	109.50	Length:	33
Ratio:	4.212	Gaps:	3
Percent Similarity:	78.788	Percent Identity:	63.636

## alignment\_block:

US-09-316-048-27 x AI614298 ..

Align seg 1/1 to: AI614298 from: 1 to: 381

```

10 gLyAlaGlyCysGlyGlyCys...CysGlyCys...gLyCysThrCysAl 24
  :::::::::::::: ::::: ::::: ::::: :::::
22 GCGTCGCGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 71
24 aGlyAlaGlyClyGlyThrCysThrAlaGlyThrGlyCysGlyAlaGly 40
  ::::: ::::: ::::: ::::: :::::
72 CGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 117
seq_name: gb_est8:AA498971

```

## seq\_documentation\_block:

```

LOCUS AA498971 399 bp mRNA EST 01-JUL-1997
DEFINITION vi83g12.r1 StrataGene mouse skin (#937313) Mus musculus cDNA clone
ACCESSION AA498971
VERSION AA498971.1 GI:2234025
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 399)

```

## REFERENCE

AUTHORS Marra, M., Hillier, L., Allen, M., Bowers, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:531102

Seq primer: -28m13 rev1 ET from Amersham  
High quality sequence stop: 379.

## FEATURES

## source

```

1. 399
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:918886"
/clone_lib="StrataGene mouse skin (#937313)"
/sex="females"
/tissue_type="whole skin"
/dev_stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skin; Vector: pBluescript SK-; Site: 1: EcoRI; Site: 2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Whole skin from 11 week old C57BL/6 female mice. Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5' CTCGAGTTT TTT TTT TTT TTT TTT 3'"
BASE COUNT 32 a 104 c 154 g 109 t
ORIGIN

```



```

/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-."
/notes="Vector: PMD42ny; Purified genomic DNA from M.
musculus C57Bl/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (g114732114[gb]/AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      62 a      82 c      206 g      135 t
ORIGIN

```

```

alignment_scores:
  Quality: 109.00      Length: 50
  Ratio: 3.303      Gaps: 2
  Percent Similarity: 66.000      Percent Identity: 38.000

```

```

alignment_block:
US-09-316-048-27 x A2434703 ..

```

```

Align seg 1/1 to: A2434703 from: 1 to: 485

```

```

1 ThrAlaCysAlaAlaThrcysThralaGlyCysGlyCysCy 17
   ::::|||||:::  ||  :::::|||||:::  ||
49 AGTGGCTGTAGCAGTGGCTGTAGCAGTGGCTGTAGCAGTGG  98
   |||||||:::  ThrcysAlaGlyAlaGlyGlyGlyThr. 30
   |||||||:::  |||
99 TGGCTGTGGCTGTGGCAGTGGCTGTGGCAGTGGCTGTGG  148
   |||||||:::  CysThralaGlyThrcysGlyAlaGly 40
   ||:::  |||:::|||||:::|||||
149 GCAGTGGCTATAGCAGTGGCTGCAGCTATGCAGTGGCTGT  198

```





OM of: US-09-316-048-27 to: N\_Geneseq\_36.\* out\_format : pfs

Date: Apr 5, 2001 8:30 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Comugen Ltd.

#### Command line parameters:

-MODEL=framed\_p2n.model -DEV=xip  
-O=/cgm2\_l/USPRO.spool/US09316048/runat\_01042001\_170039.7768/app-query.fasta.1.1293  
-DB=N\_Geneseq\_36 -QFMT=fasta -SUFFIX=p2n.rng -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blscom62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext\_MINLEN=0 -MAXLEN=200000000  
-USER=US09316048 -CGM1\_1\_248 -NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPHY  
-WAIT -THREADS=1

#### Search information block:

Query: US-09-316-048-27  
Query length: 40  
Database: N\_Geneseq\_36.\*  
Database sequences: 480022  
Database length: 187831343  
Search time (sec): 885.780000

#### score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	i	Documentation
/SID56/gcgdata/geneseq/genesqn/NA1988.DAT:V00469 -				109.50	161.66	0.3240
/SID56/gcgdata/geneseq/genesqn/NA1989.DAT:N90746 -				105.00	161.87	0.3154
/SID56/gcgdata/geneseq/genesqn/NA1999.DAT:X39707 -				104.50	163.22	0.2852
/SID56/gcgdata/geneseq/genesqn/NA1992.DAT:Q26509 -				102.50	148.56	1.74
/SID56/gcgdata/geneseq/genesqn/NA1998.DAT:V18480 -				101.50	147.64	1.96
/SID56/gcgdata/geneseq/genesqn/NA1990.DAT:Q03665 -				100.00	146.37	2.30
/SID56/gcgdata/geneseq/genesqn/NA1997.DAT:T75505 -				98.00	160.32	0.3846
/SID56/gcgdata/geneseq/genesqn/NA2000.DAT:Z49963 -				99.00	160.32	0.3846
/SID56/gcgdata/geneseq/genesqn/NA1991.DAT:Q14320 -				97.50	141.06	4.55
/SID56/gcgdata/geneseq/genesqn/NA1989.DAT:N90096 -				95.50	132.86	13.03
/SID56/gcgdata/geneseq/genesqn/NA1998.DAT:V38816 -				95.50	145.99	2.42
/SID56/gcgdata/geneseq/genesqn/NA2000.DAT:Z93361 -				94.50	149.46	1.55
/SID56/gcgdata/geneseq/genesqn/NA1998.DAT:V20700 -				94.50	132.25	14.08
/SID56/gcgdata/geneseq/genesqn/NA1998.DAT:V20701 -				94.50	132.02	14.51
/SID56/gcgdata/geneseq/genesqn/NA1997.DAT:T80415 +				94.00	123.58	42.79
/SID56/gcgdata/geneseq/genesqn/NA1997.DAT:T8508 +				94.00	114.41	138.80
/SID56/gcgdata/geneseq/genesqn/NA1997.DAT:T75504 +				93.00	154.22	0.8410
/SID56/gcgdata/geneseq/genesqn/NA1999.DAT:V73805 +				92.00	113.96	147.06
/SID56/gcgdata/geneseq/genesqn/NA1998.DAT:V19941 +				92.00	102.43	645.49
/SID56/gcgdata/geneseq/genesqn/NA1996.DAT:T30308 -				91.50	133.47	12.04
/SID56/gcgdata/geneseq/genesqn/NA1991.DAT:Q14319 -				91.50	132.07	14.41
/SID56/gcgdata/geneseq/genesqn/NA1992.DAT:Q34566 -				91.50	129.92	19.00
/SID56/gcgdata/geneseq/genesqn/NA1998.DAT:T30309 -				91.50	125.12	35.14
/SID56/gcgdata/geneseq/genesqn/NA1999.DAT:Z42235 -				90.00	135.19	9.66
/SID56/gcgdata/geneseq/genesqn/NA1994.DAT:Q73383 +				88.50	134.15	11.04
/SID56/gcgdata/geneseq/genesqn/NA1996.DAT:T44325 -				88.50	121.39	56.68
/SID56/gcgdata/geneseq/genesqn/NA1993.DAT:O51557 -				88.50	121.39	56.68
/SID56/gcgdata/geneseq/genesqn/NA1999.DAT:Z22072 +				88.50	116.31	108.79
/SID56/gcgdata/geneseq/genesqn/NA2000.DAT:Z58381 +				88.00	135.43	9.37
/SID56/gcgdata/geneseq/genesqn/NA1989.DAT:N91636 -				87.50	131.05	16.43
/SID56/gcgdata/geneseq/genesqn/NA2000.DAT:Z51983 +				86.00	135.18	9.68
/SID56/gcgdata/geneseq/genesqn/NA2000.DAT:Z45837 -				86.00	124.32	38.97
/SID56/gcgdata/geneseq/genesqn/NA1999.DAT:X60096 +				85.00	110.00	244.33
/SID56/gcgdata/geneseq/genesqn/NA1999.DAT:X33466 +				85.00	136.87	7.79
/SID56/gcgdata/geneseq/genesqn/NA1998.DAT:V03553 -				85.50	109.97	245.40
/SID56/gcgdata/geneseq/genesqn/NA2000.DAT:A31042 -				83.00	148.17	1.83
/SID56/gcgdata/geneseq/genesqn/NA2000.DAT:A00746 -				84.00	139.26	5.73

/SID56/gcgdata/geneseq/genesqn/NA1993.DAT:Q41087 -	84.00	121.27	57.59	2838
/SID56/gcgdata/geneseq/genesqn/NA1999.DAT:Z09483 -	84.00	120.54	63.26	3112
/SID56/gcgdata/geneseq/genesqn/NA2000.DAT:Z33362 -	83.50	122.53	48.97	2203
/SID56/gcgdata/geneseq/genesqn/NA1990.DAT:Q04710 +	83.00	139.91	5.27	225
/SID56/gcgdata/geneseq/genesqn/NA1996.DAT:T43976 +	83.00	139.91	5.27	225

seq\_name: /SID56/gcgdata/geneseq/genesqn/NA1998.DAT:V00469

#### seq\_documentation\_block:

ID V00469 standard; DNA; 2160 BP.

AC V00469;

DT 21-MAY-1998 (first entry)

DE E. coli zipa gene.

KW zipa gene; antimicrobial; antibiotic; screening; ss.

OS Escherichia coli.

FH Key location/Qualifiers

FT CDS 717..1703

FT sig\_peptide 717..782

FT mat\_peptide 738..1700

FT De Boer PAJ, Hale CA;

PN W09744481-A1.

PD 27-NOV-1997.

PF 21-MAY-1997; 97WO-US08703.

PR 21-MAY-1996; 96US-0651818.

PA (UYCA-) UNIV CASE WESTERN RESERVE.

PT Method for screening for antimicrobial compounds - using the zipa

PS gene of E. coli

PS Claim 12; Fig 4; 55pp; English.

This nucleotide sequence comprises the previously unidentified zipa gene located at 52 minutes on the Escherichia coli chromosome. A lambda gfl1 library made from E. coli chromosomal DNA was probed with HKT-ftsz to identify clones capable of expressing zipa. 7 Recombinant phage were identified, all of which contained 7-10 kb of E. coli DNA, including the complete 2160 bp zipa gene. The gene encodes zipa protein (see W36984) that interacts with the essential division protein FtsZ. The invention relates to screening compounds for antimicrobial activity, especially by using bacterial proteins in vitro to detect compounds that interfere with cell division. An expression vector has been constructed to allow for zipa protein to be functionally over-expressed in bacterial cells or other hosts. The zipa protein can be used in a cell-free assay to screen compounds for their antimicrobial activity, and may also be used in a cellular assay such as a yeast two-hybrid system. The antimicrobial compounds identified by the method are effective against bacteria, and may also have activity against fungi, mycoplasma and protozoa.

Sequence 2160 BP; 587 A; 546 C; 548 G; 479 T; 0 other:

alignment\_scores: 109.50

Quality: 109.50

Length: 56



CC The invention relates to a method for diagnosing a disorder characterised  
 CC by expression of a human cancer associated antigen precursor coded for by  
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
 CC biological sample isolated from a subject with an agent that specifically  
 CC binds to the NAM, an expression product or a fragment of an expression  
 CC product complexed with an HLA molecule; and (b) determining the  
 CC interaction between the agent and the NAM or the expression product as a  
 CC determination of the disorder. The products and methods can be used in  
 CC the diagnosis, monitoring, research, or treatment of conditions  
 CC characterised by the expression of various cancer associated antigens.  
 CC The invention provides nucleic acid sequences and encoded polypeptides  
 CC which are cancer associated antigen precursors expressed in human breast  
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
 CC lung cancer.

SO Sequence 691 BP; 162 A; 247 C; 139 G; 135 T; 8 other;

# alignment\_scores:

Quality: 104.50 Length: 36  
 Ratio: 4.354 Gaps: 2  
 Percent Similarity: 66.667 Percent Identity: 55.556

# alignment\_block:

US-09-316-048-27 x X39707/rev ..

Align seg 1/1 to reverse of: X39707 from: 1 to: 691

7 CysThrAlaGlyAlaGlyCysGlyGlyCys.....CysGlyCys 19  
 ||| ::|||::||| ||||| |||||  
 106 TGTGTGGGAGGGGGTGGGGTGGGGCTGAGGCTCGCGCTG 57  
 19 sGlyCysThrCysAlaGlyAlaGlyGlyGlyThrCysThrAlaGlyThr 36  
 | ||| |||::| ::||| ||||| ||| ||||| |  
 56 CTGTGTGGCGCTGCTGCTGCGGTGGAGGC...TGTGTCTGTGTGTG 10  
 36 lYcysGly 38  
 |||||  
 9 GTTGTGGC 2

seq\_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1992.DAT:Q26509

# seq\_documentation\_block:

ID Q26509 standard; DNA: 3000 BP.

AC Q26509;

XX 08-JAN-1993 (first entry)

DE Bordetella parapertussis prn gene.

KM Whooping cough; P70 antigen; P95 precursor protein; vaccination; ss.

OS Bordetella parapertussis.

XX Key Location/Qualifiers

XX CDS 145..2913

XX mat\_peptide /tag= a

XX repeat\_region /product= P95\_precursor

XX repeat\_unit /tag= b

XX repeat\_region /product= P70

XX repeat\_unit /tag= c

XX repeat\_region /tag= d

XX repeat\_unit /tag= e

XX repeat\_region /tag= f

XX repeat\_unit /tag= g

XX repeat\_region /tag= h

XX WO9211292-A.

XX 09-JUL-1992.

XX 23-DEC-1991; 91WO-GB02302.

XX 21-DEC-1990; 90GB-0027901.

XX (WELL ) WELLCOME FOUND LTD.

XX Charles IG;

XX WPI; 1992-250033/30.

XX P-PDB; R25578.

XX Acellular vaccine for immunisation against whooping cough -

XX and capable of binding antibodies which bind native P70 antigen

XX Claim 2; Fig 1; 20pp; English.

CC The DNA encodes the p95 precursor of B.parapertussis which is  
 CC processed to give the P70 antigen. A cosmid library was constructed  
 CC by transforming E.coli HB101 with recombinant cosmids prepared  
 CC by partial digestion of B.parapertussis chromosomal DNA  
 CC with Sau3A and cloning of 40-50kb fragments into the BamHI site of  
 CC cosmid pHC79. The cosmids were screened with a 1.8kb ClaI fragment  
 CC from the prn gene of B.pertussis. The insert from one positive  
 CC colony, harbouring cosmid pBD811, was sequenced and found to  
 CC contain an open reading frame encoding a 922 amino acid protein  
 CC with calculated mol.wt. 95,177.

SO Sequence 3000 BP; 451 A; 1025 C; 1102 G; 422 T; 0 other;

# alignment\_scores:

Quality: 102.50 Length: 43  
 Ratio: 3.942 Gaps: 2  
 Percent Similarity: 60.465 Percent Identity: 53.488

# alignment\_block:

US-09-316-048-27 x Q26509/rev ..

Align seg 1/1 to reverse of: Q26509 from: 1 to: 3000

4 AlaAlaThrCysThrAlaGlyAla.....GlyCysGly 14  
 |||::| ||| ||||| |||||  
 1985 GCCGCGGTGGTGGCGCGCGCTTCCGCGTCCCTGTGTGGCGCTGCGG 1936  
 14 yGlyCysCysGlyCysGlyCysThrCysAlaGlyAlaGlyGlyGlyThrC 31  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 1935 CGGCTGCGCGCGCTGCGGC...GGCTGCGCGCGCTGCGCGCTGCGGC 1889  
 31 yThrAlaGlyThrGlyCysGlyAlaGly 40  
 ||| ||| ||||| ||||| |||||  
 1888 CGGCGTGGGACCGCGGCTGCGCGCGGCG 1860

seq\_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1998.DAT:V18480

seq\_documentation\_block:

ID V18480 standard; cDNA to mRNA; 2790 BP.

XX V18480;

XX 18-AUG-1998 (first entry)

XX BOP1 cDNA.

XX Corticotroph; pituitary; BOP1; tumour suppressor; constitutive;  
 KW Inducible; Alzheimer's disease; nuclear transcription factor; apoptosis;  
 XX cell cycle; neuronal disorder; ss.

```

OS Mus sp.
XX
XX Key Location/Qualifiers
XX CDS 542..2545
XX FT /*tag=a
XX FT /product="BOP1 protein"
XX
XX W09813489-A1.
XX
XX 02-APR-1998.
XX
XX 22-SEP-1997; 97WO-EP05198.
XX
XX 23-SEP-1996; 96US-0718661.
XX
XX (CNRS ) CENT NAT RECH SCI
XX (PLAC ) MAX PLANCK GBS FOERDERUNG WISSENSCHAFTEN.
XX
XX Journot L, Spengler D;
XX
XX WPI: 1998-230701/20.
XX DR P-PSDB; W48760.
XX
XX New isolated tumour suppressor gene - useful for developing products
XX PT for use in diagnosis and treatment of tumour(s) or neuronal
XX PT disorder(s)
XX
XX PS Claim 1: Pages 72-76; 118pp; English.
XX
XX The present sequence represents the BOP1 cDNA isolated from the
XX CC mouse corticotroph pituitary tumour cell line AtT-20 cDNA library.
XX CC The protein encoded by the BOP1 cDNA displays a tumour suppressing
XX CC activity when it was constitutively and inducibly expressed in
XX CC tumour cells. The BOP1 cDNA and the protein it encodes are claimed
XX CC to be useful in the preparation of therapeutic compositions, useful
XX CC for treating, preventing or delaying the recurrence of a tumour or
XX CC neuronal disorders, e.g.: genetic diseases or acquired degenerative
XX CC encephalopathies such as Alzheimer's disease. The BOP1 protein is
XX CC also claimed to be able to induce apoptosis resulting in inhibition
XX CC of tumour cell growth, to suppress tumour formation, to induce G1
XX CC arrest of the cell cycle and to act as nuclear transcription factor.
XX
XX SQ Sequence 2790 BP; 667 A; 783 C; 714 G; 626 T; 0 other;

```

alignment\_scores:

Quality:	101.50	Length:	47
Ratio:	2.985	Gaps:	4
Percent Similarity:	72.340	Percent Identity:	42.553

alignment\_block:

```

US-09-316-048-27 x V18480/rev ..

```

Align seg 1/1 to reverse of: V18480 from: 1 to: 2790

```

1 ThrAlaCysAlaAlaThrCysThrAlaGlyAlaGlyCysGlyCys.. 16
:::|||||::: |||:::|||||:::|||||:::|||||
2286 TCAGGCTGCGCTGAGGTTGACATTGCGCAGCAGCGTGTGATGCAT 2237
17 .....Cys.....GlyCysGlyCysThrCysAlaGlyAlaG 27
:::|||||::: |||:::|||||::: |||::: |||::: |||
2236 CTGGGCTGCACTGCAACTGAACTGTGCTCCTCATCTGT...GGCTGTG 2190

```

27 lyGlyGlyThrCysThrAlaGlyAlaGlyCysGlyAlaGly 40  
 ||:::||||| |||:::|||||:::|||||:::|||||  
 2189 GTAGTGGC...TGTGCACCTGGCAGTGGCTGTGGCAGTGGC 2152

seq\_name: /SID6/gcgdata/geneseq/geneseq/NA1990.DAT:Q03665

seq\_documentation\_block:  
 ID Q03665 standard; DNA: 2465 BP.  
 XX  
 AC Q03665;

```

XX
XX 07-AUG-1989 (first entry)
XX
XX Sequence homologous to Drosophila Per gene.
XX DE
XX Sex determination; ruminant embryos; ss.
XX
XX FR2635116-A.
XX
XX 09-FEB-1990.
XX
XX 08-AUG-1988; 88FR-0010706.
XX
XX 08-AUG-1988; 88FR-0010706.
XX
XX (GEOR/) GEORGES M.
XX
XX Georges M, Vassart G, Christophe D, Dumont J, Young M;
XX PI
XX DR WPI: 1990-093373/13.
XX
XX Determn. of sex of ruminant embryos-by hybridisation assay using specific
XX PT probes.
XX
XX PS Claim 17; fig 1; 37pp; French.
XX
XX This sequence is used as a probe and is contained in plasmid pSP64.2 SEI.
XX CC It is homologous to the sequence of a 2.5 kb fragment of the Per gene of
XX CC Drosophila. The motifs ACNCGN and/or TCAAGC (N=T, U, G, A, or C) or their
XX CC complementary sequences are repeated in the probe. It is used in a
XX CC hybridisation assay to determine the sex of ruminant embryos. Either
XX CC (i) Southern- or (ii) Dot-Plot techniques can be utilised and Y-chromo-
XX CC some shows up as intense signal either obscuring most of the hybridisa-
XX CC ion track in (i) or surrounding the dots in (ii). Significant results can
XX CC be obt. using samples of less than 50 ng weight.
XX
XX SQ Sequence 2465 BP; 791 A; 739 C; 689 G; 246 T; 0 other;

```

alignment\_scores:

Quality:	100.00	Length:	36
Ratio:	4.000	Gaps:	0
Percent Similarity:	69.444	Percent Identity:	44.444

alignment\_block:

```

US-09-316-048-27 x Q03665/rev ..

```

Align seg 1/1 to reverse of: Q03665 from: 1 to: 2465

```

2 AlaCysAlaAlaThrCysThrAlaGlyAlaGlyCysGlyCysG 18
|||||:::|||||:::|||||:::|||||
593 GCCTGTGCCGTGCTGCTGTGCTGCTGCTGCTGCTGCTGCTG 544
18 yCysGlyCysThrCysAlaGlyAlaGlyGlyGlyThrCysThrAlaGlyT 35
|||||:::|||||::: |||::: |||::: |||
543 CTGTGCTGTCTCTCTGTTCCGTGCTGCTGCTGCTGCTGCTG 494
35 hrGlyCys 37
:::|||||
493 CTGCTGT 486

```

seq\_name: /SID6/gcgdata/geneseq/geneseq/NA1997.DAT:T75505

seq\_documentation\_block:  
 ID T75505 standard; cDNA: 292 BP.  
 XX  
 AC T75505;  
 XX  
 XX 23-MAR-1998 (first entry)  
 XX  
 DE P. americanus skin type AFP, clone sAF8 cDNA.  
 XX  
 XX Antifreeze polypeptide; intracellular skin type AFP; winter flounder;

```

KW cold resistance; thermal hysteresis; antibacterial; ss.
XX
OS Pseudopleuronectes americanus.
XX
Key Location/Qualifiers
FH 36..200
CDS /*lag- a
FT /product= sAFP8
FT /note= "skin type antifreeze polypeptide"
XX
PN M09728260-A1.
PD
XX
07-AUG-1997.
XX
PE 30-JAN-1997; 97MO-CA00062.
XX
PR 31-JAN-1996; 96US-0010920.
XX
PA (HSCR-) HSC RES & DEV LP.
PI
XX Gong Z, Hew C;
XX
DR WPJ: 1997-402614/37.
DR P-PSDB: W22875.
XX
PT Isolated fish skin anti-freeze polypeptide - useful to depress
PT freezing points of aqueous compositions and protect plant, fungal,
XX animal or bacterial cells from cold
XX
PS Claim 11; Figure 11; 104pp; English.
XX
CC This sequence encodes the sAFP8 clone corresponding to a novel skin type
CC intracellular antifreeze polypeptide (AFP) which induces a concentration
CC dependent decrease in the freezing point of an aqueous solution. This
CC novel polypeptide does not contain a signal sequence and is thought to be
CC intracellular. AFP's can be used to make an aqueous composition resistant
CC to freezing by changing its thermal hysteresis such as a water or salt
CC solution, an intracellular compartment of a cell or a food stuff, e.g.
CC soft serve 'frozen' yogurt or ice cream. AFP's can inhibit ice
CC recrystallisation during cold storage, improving the texture and
CC palatability of the food and has antibacterial properties. Such
CC polypeptides can also be expressed to provide cold resistance to cells,
CC e.g. plant, fungal animal or bacterial cells. The antibodies can be used
CC to identify and isolate AFP while its promoter can be used to direct
CC expression of a nucleic acid.
XX
SQ Sequence 292 BP; 66 A; 109 G; 76 G; 41 T; 0 other;
XX
alignment_scores:
Quality: 98.00 Length: 33
Ratio: 4.261 Gaps: 3
Percent Similarity: 69.697 Percent Identity: 60.606
XX
alignment_block:
US-09-316-048-27 x T75505/rev ..
XX
Align seg 1/1 to reverse of: T75505 from: 1 to: 292
XX
10 gIyAlaGlyCySgLyGlyCys...CysgLy..CysgLyCysThrCysAI 24
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
137 GGcAGtGGcTGGCGGCGCTGCgcGTGCGGCGTGcGTGcCTTCVCgG 88
24 aGIyAlaGIyGlyGlyThrCysThrAlaGIyThhGlyCySgLyAlaGIy 40
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
87 CGGcCTTGcGCGGCGGC.....GGCGGTGcGTGcGTGcGCGGCGGC 51
XX
seq_name: /SID56/gcgdata/geneseq/geneseqn/NA2000.DAT.Z49963
seq_documentation_block:
ID Z49963 standard; cDNA; 292 BP.
XX
Z49963;
AC
```

```

XX 25-APR-2000 (first entry)
XX
XX Winter flounder skin-type antifreeze protein-8 DNA.
DE
XX Winter flounder skin-type antifreeze protein; wfsAFP-8; antibacterial;
KW cold tolerance; ice recrystallisation; refrigerated food;
KW antifreeze; ds.
XX
OS Pseudopleuronectes americanus.
XX
XX MO200000512-A2.
XX
XX 06-JAN-2000.
XX
XX 25-JUN-1999; 99MO-CA00601.
XX
XX 26-JUN-1998; 98US-0090794.
PR 07-AUG-1998; 98US-0095713.
PR 24-JUN-1999; 99US-0344529.
XX
XX (HSCR-) HSC RES & DEV LP.
PA
PI Hew CL;
XX
XX WPI: 2000-170905/15.
XX
XX Novel antifreeze polypeptides and polynucleotides used to make cells
PR cold resistant and to improve the palatability of cold foods and
PT liquids -
XX
XX Disclosure; Fig 5; 61pp; English.
PS
XX
XX The present sequence encodes winter flounder skin type antifreeze
CC protein (wfsAFP-8). This lacks a signal peptide and is indicative of its
CC intracellular location. Polyclonal antisera generated using scupin skin
CC type AFP-2 is tested for its cross-reactivity against wfsAFPs using a
CC competitive binding immunoassay. The APFs are used to make aqueous
CC compositions resistant to freezing, to improve the cold tolerance, for
CC inhibiting ice recrystallisation, and for stabilising biological
CC membranes. They are also used to improve the shelf life and food quality
CC of many refrigerated foods and have antibacterial properties.
XX
XX Sequence 292 BP; 66 A; 109 C; 76 G; 41 T; 0 other.
SQ

alignment_scores:
    Quality: 98.00      Length: 33
    Ratio: 4.261      Gaps: 3
    Percent Similarity: 69.697      Percent Identity: 60.606

alignment_block:
US-09-316-048-27 x Z49963/rev ..

Align seg 1/1 to reverse of: Z49963 from: 1 to: 292

    10 GYALAGAGLYCysGlyGlyCys...CysGly...CysGlyCysThrCysA1 24
       |||:::||||| ||||| ||||| |||||
    137 GCGAGGCTGCGGCGGCGCTCGCGCTGCGGCACTGCGGCTCTCTGCGG 88

    24 aGlyAAGAGlyGlyGlyThrCysThrAAGAGlyThrGlyCysGlyAAGly 40
       :||| ||||| ||||| ||||| ||||| |||||
    87 CGGCGCTTGCGCGCGGC.....GGCGGTGCGTCTGCGGCGGCG 51

seq_name: /SIS06/gcgdata/geneseq/geneseqn/NA1991.DAT.Q14320
seq_document_block:
ID Q14320 standard; DNA; 3000 BP.
AC Q14320;
XX
XX 20-JAN-1992 (first entry)
XX

```

DE	Pertactin antigen P.70.
XX	
KW	Pertactin; Pichia; B. pertussis; B.bronchiseptica; ss.
XX	
OS	Bordetella paraperussis.
XX	
FT	Key
FT	Location/Qualifiers
FT	CDS 145..2910
FT	/*tag= a
FT	/label= pertactin
XX	
PN	W09115571-A.
XX	
PD	17-OCT-1991.
XX	
PE	28-MAR-1991; 91MO-GB00487.
XX	
PR	02-APR-1990; 90GB-0007416.
XX	
PA	(WELT ) WELLCOME FOUNDATION LTD.
XX	
PI	Clare JJ, Romanos MA;
XX	
DR	WPI; 1991-325214/44.
DR	P-PSDB; RI4321.
XX	
PT	Pichia microorganism transformants - for production of
PT	Bordetella pertactin antigens for whooping cough vaccines
XX	
SS	Disclosure; Fig 1C; 38pp; English.
XX	
CC	Pichia microorganisms are transformed for the expression of
CC	pertactin antigens. DNA sequence used are represented in Q14319-20
CC	encoding the B. bronchiseptica P.68 and B. paraperussis P.70 antigen
CC	respectively or the B. pertussis P.69 encoding sequence described
CC	by I.G. Charles et al. Proc. Natl. Acad. Sci. USA, Vol. 80:3554-3448
CC	(1989).
CC	The 109 Ns represent nucleotides missing in the specification.
XX	
Q0	Sequence 3000 BP; 427 A; 987 C; 1073 G; 404 T; 109 other;

```

alignment_scores:
    Quality: 97.50      Length: 42
    Ratio: 3.611      Gaps: 2
    Percent Similarity: 64.286      Percent Identity: 52.381

alignment_block:
US-09-316-048-27 x Q14320/rev ..

Align seg 1/1 to reverse of: Q14320 from: 1 to: 3000

      4 AlaAlaThrCysThrAlaGlyAlaGly.....CysGlyGlyCysCy 17
      ||::: ||| |||||:::|::: |||||:::|
1985 GCGGGGGGTTGGCGGGCGCGCGCTTCGCGCTCGTGTGGCGGCTCGCG 1936

      17 sGlyCysGlyCysThr.....CysAlaGlyAlaGlyGlyGlyThrCyst 32
      |||||:::|::: |||||:::|:::|:::|:::|:::|
1935 CGGCTTCGCGGNNNNNCGCGCGCTCGCGCGCTCGCGGCTGGGGACCGG 1886

      32 hrAlaGlyThrGlyGlyCysGlyAlaGly 40
      ||| |||||:::|:::|:::|:::|:::|
1885 GCTGGGGACCGGCGCTCGCGCGCGGC 1860

seq_name: /SID56/gcgdata/geneseq/geneseqn/NA1989.DAT:N90096
seq_documentation_block:
ID N90096 standard, cDNA, 6978 BP.
XX
AC N90096:
XX
DT 22-MAR-1991 (first entry)
XX

```

DE	Sequence of plasmid pAY31 encoding wheat alpha-gliadin.
XX	
KW	Coeliac disease diagnosis; dough formation; food technology; ds.
XX	
OS	Wheat.
XX	
PN	USA826765-A.
XX	
PD	02-MAY-1989.
XX	
PF	24-OCT-1986; 86US-0922616.
XX	
PR	24-OCT-1986; 86US-0922616.
XX	
PA	(UYHA-) UNIV OF HAWAII (USDA).
XX	
PI	Greene FC, Stiles JT, Neill JD, Anderson OD, Litts JC;
DR	WPI; 1989-150270/20.
XX	
XX	
PT	New wheat gluten protein gene plasmid - useful for genetic
XX	transformation of yeasts to produce gliadin or glutenin
PS	Disclosure; Fig 4; 21pp; English.
XX	
CC	The preferred gene fragment in the patent of the invention encodes
CC	alpha-gliadin or glutenin, and the preferred plasmid encoding alpha-
CC	gliadin is PAY 31. The plasmid may be used for the prodn. of alpha-
CC	gliadin, which may be used in the food technology industry to modify
CC	prodn. processes of doughs and batters and to reduce unit costs.
CC	Gliadin may also be used for diagnosis and treatment of illness
CC	caused by wheat gluten proteins and for testing theories of dough
CC	formation.
SQ	Sequence 6978 BP; 2042 A; 1631 C; 1447 G; 1858 T; 0 other;

```

alignment_scores:          Quality: 96.50          Length: 43
                          Ratio: 3.860          Gaps: 2
Percent Similarity: 58.140          Percent Identity: 44.186

alignment_block:
us-09-316-048-27 x N90096/rev ..

Align seg 1/1 to reverse of: N90096 from: 1 to: 6978

3 CysAlaIaIaThrcysTrhAlaGlAlaGlCysGlyCly..... 15
||||:||||: ||| :||| |||
3080 TGGTGGCTGCGTGGCTGCTGCGAATATGGTTGTGGTTCGATACTG 3031

16 CysGlyGlyCysGlyCysThrcysAlaGlAlaGlGlyGlyThrcysT 32
|||||:|||||: |||:|||||: |||:|||||: |||
3030 TTGTGGGGGTGGATGATGATGTTGTGTGCGAATATGGTTGTGGCTGTG 2981

32 hrAlaGlyThr...GlyCysGlyAlaGly 40
||||:||||: |||:|||||: |||
2980 AATATGCTAGTTCGCGCTTCGCGGAATATGCT 2952

seq_name: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT.V38816
seq_documentation_block:
ID V38816 standard; DNA; 1107 BP.
XX
AC V38816;
XX
XX
DT 09-OCT-1998 (first entry)
XX
DE DNA encoding a durum wheat glutenin protein.
XX
KW Glutenin gene; durum wheat; low-molecular-weight;
KW transgenic durum wheat; ss.
XX

```



```
XX XX Cryptosporidium parvum GP900 antigen open reading frame.
XX DE
XX XX Open reading frame; ORF; antigen: GP900; cryptosporidium: infection;
XX KW antibody; prophylaxis; treatment; inhibition; retardation;
XX KW detection; diagnosis; human; ds.
XX OS Cryptosporidium parvum.
XX FH
XX FH Key Location/Qualifiers
XX CDS 1..5163
XX FT /*tag= a
XX FT /product= GP900 antigen
XX FT misc_feature 524..1270
XX FT /*tag= b
XX FT /note= "region containing NINC mutations"
XX FT
XX PN WO9806430-A1.
XX PD 19-FEB-1998.
XX PF 11-AUG-1997; 97WO-US14104.
XX PR 14-AUG-1996; 96US-0700651.
XX PA (REGC ) UNIV CALIFORNIA.
XX PI Gut J, Leech J, Nelson RC, Petersen C;
XX DR WPI: 1998-159290/14.
XX DR P-PSDB: W48299.
XX PT Anti-Cryptosporidium antibody - used to develop products for
XX PT detection, diagnosis, prophylaxis or treatment of Cryptosporidium
XX PT infections
XX PS Claim 32; Pages 60-62; 89pp; English.
XX CC The sequence is that encoding the GP900 antigen which may be used
XX CC in the production of anti-cryptosporidium antibodies. These can be
XX CC used for the prophylaxis, treatment, inhibition or retardation of
XX CC a Cryptosporidium infection in humans or in animals such as calves.
XX CC They can also be used for the detection and diagnosis of related
XX CC infections.
XX SQ Sequence 5163 BP; 1875 A; 1137 C; 873 G; 1278 T; 0 other;

alignment_scores:
    Quality: 94.50      Length: 31
    Ratio: 4.109      Gaps: 1
    Percent Similarity: 74.194      Percent Identity: 45.161

alignment_block:
US-09-316-048-27 x V20700/rev ..
Align seg 1/1 to reverse of: V20700 from: 1 to: 5163

7 CysThAlaGlyAlaGlyCysGlyGlyCysGlyCysGlyCysThrcy 23
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
865 TGTAGTTGCGTGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 816
23 sAlaGlyAlaGlyGlyGlyThrcysThraGlyThrcGlyCys 37
|||:|||||:|||||:|||||:|||||:|||||:|||||:
815 TTGTGCT.....TGTAGTTGCTGCTGCTGCTGCTGCTGCTG 788

seq_name: /SID56/gcdata/geneseq/geneseqn/NA1998.DAT:V20701
seq_documentation_block:
ID V20701 standard; DNA: 5318 BP.
XX AC
XX V20701;
XX
```

```
DT 17-AUG-1998 (first entry)
XX DE
XX XX Cryptosporidium parvum GP900 antigen open reading frame and 3' region.
XX KW Open reading frame; ORF; antigen: GP900; cryptosporidium: infection;
XX KW antibody; prophylaxis; treatment; inhibition; retardation;
XX KW detection; diagnosis; human; 3' region; ds.
XX OS Cryptosporidium parvum.
XX FH
XX FH Key Location/Qualifiers
XX CDS 1..5166
XX FT /*tag= a
XX FT /product= GP900 antigen
XX FT misc_feature 524..1270
XX FT /*tag= b
XX FT /note= "region containing NINC mutations"
XX FT
XX PN WO9806430-A1.
XX PD 19-FEB-1998.
XX PF 11-AUG-1997; 97WO-US14104.
XX PR 14-AUG-1996; 96US-0700651.
XX PA (REGC ) UNIV CALIFORNIA.
XX PI Gut J, Leech J, Nelson RC, Petersen C;
XX DR WPI: 1998-159290/14.
XX DR P-PSDB: W48299.
XX PT Anti-Cryptosporidium antibody - used to develop products for
XX PT detection, diagnosis, prophylaxis or treatment of Cryptosporidium
XX PT infections
XX PS Claim 32; Pages 62-63; 89pp; English.
XX CC The sequence is that encoding the GP900 antigen which may be used
XX CC in the production of anti-cryptosporidium antibodies. These can be
XX CC used for the prophylaxis, treatment, inhibition or retardation of
XX CC a Cryptosporidium infection in humans or in animals such as calves.
XX CC They can also be used for the detection and diagnosis of related
XX CC infections.
XX SQ Sequence 5318 BP; 1944 A; 1153 C; 888 G; 1333 T; 0 other;

alignment_scores:
    Quality: 94.50      Length: 31
    Ratio: 4.109      Gaps: 1
    Percent Similarity: 74.194      Percent Identity: 45.161

alignment_block:
US-09-316-048-27 x V20701/rev ..
Align seg 1/1 to reverse of: V20701 from: 1 to: 5318

7 CysThAlaGlyAlaGlyCysGlyGlyCysGlyCysGlyCysThrcy 23
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
865 TGTAGTTGCGTGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 816
23 sAlaGlyAlaGlyGlyGlyThrcysThraGlyThrcGlyCys 37
|||:|||||:|||||:|||||:|||||:|||||:|||||:
815 TTGTGCT.....TGTAGTTGCTGCTGCTGCTGCTGCTGCTG 788

seq_name: /SID56/gcdata/geneseq/geneseqn/NA1997.DAT:T80415
seq_documentation_block:
ID T80415 standard; DNA: 13987 BP.
XX AC
XX T80415;
XX
```







OM of: US-09-316-048-27 to: GenEmbl.\* out\_format : pfs  
Date: Apr 5, 2001 10:40 AM

About: Results were produced by the Gencore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

## Command line parameters:

-MODEL=ifame+p2n.model -DEV=xlp  
-O=/sgn2.1/USPRO.spool/US09316048/runat.01042001.170039\_7742/app-query.fasta\_1.1293  
-DB=genEmbl -OEMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPOL=0.000 -LOOPEXT=0.000  
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPEXT=10.000 -YGAPEXT=0.500  
-DELDP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NOR=ext -MINLEN=0 -MAXLEN=200000000  
-USER=US09316048 -ECGN1\_1.9054 -NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPXY  
-WAIT -THREADS=1

## Search information block:

Query: US-09-316-048-27  
Query length: 40  
Database: GenEmbl.\*  
Database sequences: 1118133  
Database length: -1736092196  
Search time (sec): 4765.960000

## score\_list:

Sequence	Strd Orig	ZScore	Escore	len	Documentation
gb_in2:BMOCHRC	129.00	194.96	0.0106	449	J01024 bombyx mori (silkmoth) c
gb_pr6:HS402G11	124.00	150.73	3.08	177241	I AL022328 human DNA sequence
gb_in2:BMOCHRC	122.50	175.95	0.1211	2182	I K02835 B.mori (silkmoth) h1g1
gb_pl3:SS1132828	120.00	172.87	0.1797	2028	I A132828 Spematozoopsis similis
gb_hlg18:AC079423	119.50	141.67	9.83	273658	I AC079423 Mus musculus chrom
gb_hlg24:DMFCHR12_07	119.00	146.63	5.20	110000	I Contination (8 of 23) of l
gb_hlg17:AC073826	118.00	143.45	7.82	146190	I AC073826 Mus musculus clone
gb_hlg17:AC073748	116.00	139.49	12.99	175371	I AC073748 Mus musculus clone
gb_hlg17:AC073748	116.00	139.49	12.99	175371	I AC073748 Mus musculus clone
gb_hlg1:AC015041	115.00	137.81	16.13	183182	I AC015041 Homo sapiens chrom
gb_ba2:BBR250095	112.50	173.70	0.1617	323	I AJ250095 Bordetella bronchisept
gb_in2:BMOCHRC	112.50	172.09	0.1986	418	I X15513 silkmoth mRNA for trans
gb_in2:BMOCHRC	112.50	163.78	0.5765	1591	I X13860 silkmoth chorion high c
gb_v12:OPR39145	112.50	156.20	1.52	5388	I U39145 Oryza pseudoturgata nu
gb_v12:OPR39145	112.50	136.31	19.53	131993	I U75930 Oryza pseudoturgata
gb_hlg10:AC023834	112.00	133.29	28.80	191692	I AC023834 Mus musculus clone
gb_in2:BMOCHRC	111.00	162.51	0.6785	1388	I X01068 Bombyx mori Hc-B.13 gert
gb_hlg5:AC014402	111.00	143.56	7.71	29275	I AC014402 Drosophila melanog
gb_in1:AE003434	111.00	129.03	49.70	302786	I AE003434 Drosophila melanog
gb_pr7:HS447742	110.50	151.02	2.96	7869	I U47742 Human monocytic leukem
gb_hlg19:AF295015	110.50	141.22	10.42	38078	I AF295015 Homo sapiens chrom
gb_hlg9:AC023045	110.00	130.78	39.73	204117	I AC023045 Homo sapiens chrom
gb_paf1:AB6713	110.00	161.90	0.7338	1220	I AB6713 Sequence 59 from patent
gb_pl3:TAGL1A1	110.00	159.92	0.9466	1679	I X02538 wheat gene for alpha/b
gb_pl3:WHG1A1B	110.00	156.22	1.52	3043	I K03074 wheat (T.aestivum) alp
gb_pl3:WHG1A1B	110.00	155.95	1.57	3178	I M16496 Trilicium urartu alpha/
gb_pl3:WHG1A1B	110.00	155.95	1.57	3178	I M16496 Trilicium urartu alpha/
gb_ba2:ECU74550	109.50	159.01	1.06	1733	I U16457 Escherichia coli CysZ
gb_paf1:AE007236	109.50	157.64	1.27	2160	I AR073236 Sequence 1 from paten
gb_ba1:AE000329	109.50	147.35	4.74	11313	I AE000329 Escherichia coli K-1
gb_ba2:DN9870	109.50	147.25	4.81	11500	I D9870 E.coli genomic DNA, KC
gb_hlg13:AC034248	109.50	130.88	39.23	160030	I AC034248 Homo sapiens chrom
gb_hlg17:AC073547	109.50	129.40	47.39	202859	I AC073547 Homo sapiens chrom
gb_ba1:AE004006	109.00	146.55	5.25	11476	I AE004006 Xylella fastidiosa,
gb_ro:MMFH4	108.00	157.73	1.25	1515	I X92591 M.musculus mRNA for fth
gb_hlg6:AC017444	108.00	135.80	20.86	51531	I AC017444 Drosophila melanog
gb_pr1:AC003958	107.00	128.74	51.59	127834	I AC003958 Homo sapiens chrom
gb_in3:PFACSPA	106.00	157.49	1.29	999	I M28887 Plasmodium berghei (clon
gb_in3:PFACSPA	106.00	157.22	1.34	1044	I X17606 Plasmodium berghei gene
gb_in3:PFACSPA	106.00	156.82	1.41	1114	I M53445 P.berghel circumsporoz

gb_in3:PFACSPA	-	106.00	156.79	1.41	1118	I M1435 Plasmodium berghei c
gb_in2:CE01B7	-	106.00	137.19	17.46	26180	I Z66499 Caenorhabditis eleg
gb_in1:AE003612	-	106.00	122.90	109.09	260550	I AE003612 Drosophila melan
gb_paf1:109305	-	105.00	156.73	1.42	900	I 109305 Sequence 1 from Paten
gb_hlg15:AC064853	+	105.00	125.67	76.51	133031	I AC064853 Homo sapiens chr

seq\_name: gb\_in2:BMOCHRC

## seq\_documentation block:

LOCUS	BMOCHRC	449 bp	mRNA	INV	26-APR-1993
DEFINITION	bombyx mori (silkmoth)	chorion hc protein mRNA.			
ACCESSION	J01024				
VERSION	J01024.1	GI:155984			
KEYWORDS	chorion protein.				
SOURCE	bombyx mori (silkmoth).				
ORGANISM	Bombyx mori				

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;  
Ditrysia; Bombycoidea; Bombycidae; Bombyx.

## REFERENCE

1 (bases 1 to 449)  
Rodakis, G.C. and Kafatos, F.C.  
origin of evolutionary novelty in proteins: how a high-cysteine

## AUTHORS

chorion protein has evolved

Proc. Natl. Acad. Sci. U.S.A. 79, 3551-3555 (1982)

## JOURNAL

## MEDLINE

## FEATURES

source 1. 449  
Location/Qualifiers  
/organism="Bombyx mori"  
/db\_xref="taxon:7091"

BASE COUNT 73 a 85 c 159 g 132 t  
ORIGIN

## alignment\_scores:

Quality:	129.00	Length:	47
Ratio:	4.031	Gaps:	2
Percent Similarity:	68.005	Percent Identity:	51.064

## alignment block:

US-09-316-048-27 x BMOCHRC

Align seg 1/1 to: BMOCHRC from: 1 to: 449

2 AlacysAlaAlaThrcysThrala.....GLYAlaG1 12

12 YcyscylgYcyscysgYcysgYcysThrcysAlaAlaGlyG 29

239 GCCTGTGCTAGTGGCTGTCTCTATTCGCGAGCTTCTGCGATCGCG 288

289 GTGCGAGAGTGGCGGTGCGATGCGGTGCGAGCGTGGTGGAT 338

29 lYThrcysThralaGly.....ThrgYcysGlyAlaGly 40

339 GTGATGCGCGCTGAGAGCTGCGAGATGCGATGCGGGA 379

seq\_name: gb\_pr6:HS402G11

## seq\_documentation block:

LOCUS	HS402G11	177241 bp	DNA	PR1	12-DEC-1999
DEFINITION	Human DNA sequence from clone RP3-402G11 on chromosome				
	22q13.31-13.33 Contains the MAPK12 gene for mitogen activated				
	protein kinase 12 (SAPK3), the MAPK11 gene for mitogen activated				
	protein kinase 11 (PRK11), the gene for a novel				
	protein similar to KIAA0501 and mouse histone deacetylase MHD2,				
	the gene for a novel protein similar to Xenopus gamma-tubulin				
	interacting protein (yeast Spc98 homolog), the gene for a novel				
	protein similar to yeast and bacterial predicted proteins, the gene				
	for a novel protein similar to C. elegans F38A5.2, the gene for a				
	novel protein similar to MRS10 and the gene for a novel protein				
	similar to mouse MOV10 (GB110) and yeast and plant predicted				
	proteins. Contains ESTs, GSSs and fifteen putative CpG islands,				
	complete sequence.				

ACCESSION

AL022328

VERSION AL022328.21 GI:5263010  
KEYWORDS HTG; CPG Island; GB10; KIAA0315; KIAA0901; kinase; MAPK11; MAPK12; MHD2; mitogen activated protein kinase; mitogen-activated; MOV10; MRS1; PRK11; SAK3; SPC98.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
AUTHORS 1 (bases 1 to 177241)  
TITLE Coville G.  
JOURNAL Direct Submission  
Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
On Jun 29, 1999 this sequence version replaced gi:5262835.  
requests: clonerequest@sanger.ac.uk  
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.  
This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP3-402G11 is from the library RPCI-3 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://biopac.med.buffalo.edu/VECTOR.pctpac2>  
This sequence is the entire insert of clone RP3-402G11 The true right end of clone RP5-89814 is at 23403 in this sequence. The true right end of clone RP11-232E17 is at 56734 in this sequence. The true right end of clone RP4-600024 is at 96013 in this sequence.  
FEATURES  
Location/Qualifiers  
source  
1..177241  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="22"  
/map="q13.31-13.33"  
/clone="RP3-402G11"  
/clone\_lib="RPCI-3"  
repeat\_region  
135..447  
/note="AluX repeat: matches 1..312 of consensus"  
repeat\_region  
657..959  
/note="AluSg repeat: matches 1..310 of consensus"  
repeat\_region  
1373..1659  
/note="AluX repeat: matches 1..298 of consensus"  
repeat\_region  
1669..1966  
/note="AluSg repeat: matches 1..306 of consensus"  
repeat\_region  
2504..2574  
/note="MER58 repeat: matches 170..243 of consensus"  
repeat\_region  
2663..2774  
/note="L2 repeat: matches 2643..2747 of consensus"  
repeat\_region  
2775..3072  
/note="AluSp repeat: matches 1..300 of consensus"  
repeat\_region  
3073..3086  
/note="L2 repeat: matches 2630..2643 of consensus"  
repeat\_region  
3959..4455  
/note="MER1A repeat: matches 31..527 of consensus"  
repeat\_region  
5044..5270  
/note="Aluub repeat: matches 77..303 of consensus"  
repeat\_region  
6305..6592

misc\_feature  
6406..6907  
/note="match: GSS: Em:A0475194"  
repeat\_region  
6853..7155  
/note="Alu repeat: matches 1..303 of consensus"  
repeat\_region  
7218..7516  
/note="AluY8 repeat: matches 1..318 of consensus"  
repeat\_region  
9018..9309  
/note="AluX repeat: matches 1..294 of consensus"  
repeat\_region  
9427..9542  
/note="AluO repeat: matches 30..150 of consensus"  
repeat\_region  
9543..9847  
/note="AluY repeat: matches 1..303 of consensus"  
repeat\_region  
9848..10020  
/note="AluO repeat: matches 150..306 of consensus"  
repeat\_region  
10024..10316  
/note="AluSg repeat: matches 1..293 of consensus"  
repeat\_region  
10324..10619  
/note="AluY repeat: matches 2..297 of consensus"  
repeat\_region  
10804..11101  
/note="AluSg repeat: matches 1..296 of consensus"  
repeat\_region  
11227..11641  
/note="MER4C repeat: matches 22..461 of consensus"  
repeat\_region  
11642..11957  
/note="AluY repeat: matches 1..310 of consensus"  
repeat\_region  
11958..11981  
/note="MER4C repeat: matches 1..22 of consensus"  
repeat\_region  
14063..14370  
/note="AluY repeat: matches 2..310 of consensus"  
repeat\_region  
15488..15718  
/note="AluX repeat: matches 51..295 of consensus"  
repeat\_region  
16411..16616  
/note="AluY repeat: matches 91..296 of consensus"  
repeat\_region  
16816..17110  
/note="Aluub repeat: matches 1..305 of consensus"  
repeat\_region  
18096..18300  
/note="MER46C repeat: matches 113..337 of consensus"  
repeat\_region  
18366..18558  
/note="MIR repeat: matches 32..247 of consensus"  
repeat\_region  
18590..18727  
/note="LTR45 repeat: matches 389..525 of consensus"  
repeat\_region  
21328..21889  
/note="MER54B repeat: matches 1..638 of consensus"  
repeat\_region  
21958..22115  
/note="L1 repeat: matches 4613..4770 of consensus"  
repeat\_region  
22116..22245  
/note="FLAMC repeat: matches 1..129 of consensus"  
repeat\_region  
22246..22887  
/note="L1 repeat: matches 3975..4613 of consensus"  
repeat\_region  
22938..23151  
/note="FRAM repeat: matches 1..166 of consensus"  
repeat\_region  
23152..23217  
/note="L1 repeat: matches 3919..3994 of consensus"  
repeat\_region  
23218..23373  
/note="AluSg/x repeat: matches 133..302 of consensus"  
repeat\_region  
23375..23514  
/note="L1 repeat: matches 3780..3926 of consensus"  
repeat\_region  
23516..23691  
/note="FRAM repeat: matches -7..168 of consensus"  
repeat\_region  
23699..24482  
/note="L1 repeat: matches 2915..3769 of consensus"  
repeat\_region  
24501..24747  
/note="LIMEC repeat: matches 1501..1751 of consensus"  
repeat\_region  
24978..25194  
/note="LIMEC repeat: matches 272..492 of consensus"  
repeat\_region  
25933..26246  
/note="AluY repeat: matches 1..309 of consensus"  
repeat\_region  
27926..28225  
/note="AluX repeat: matches 2..302 of consensus"  
repeat\_region  
28694..28997  
/note="AluSg repeat: matches 1..304 of consensus"  
repeat\_region  
29001..29119  
/note="AluO/FLAM repeat: matches 3..133 of consensus"





```

* 13907 14006: gap of unknown length
* 14007 15090: contig of 1084 bp in length
* 15091 15190: gap of unknown length
* 15191 16431: contig of 1241 bp in length
* 16432 16531: gap of unknown length
* 16532 17756: contig of 1225 bp in length
* 17757 17856: gap of unknown length
* 17857 19038: contig of 1182 bp in length
* 19039 19138: gap of unknown length
* 19139 20372: contig of 1234 bp in length
* 20373 20472: gap of unknown length
* 20473 21712: contig of 1240 bp in length
* 21713 21812: gap of unknown length
* 21813 22979: contig of 1167 bp in length
* 22980 23079: gap of unknown length
* 23080 24142: contig of 1063 bp in length
* 24143 24242: gap of unknown length
* 24243 25417: contig of 1175 bp in length
* 25418 25517: gap of unknown length
* 25519 26565: contig of 1048 bp in length
* 26566 27713: contig of 1048 bp in length
* 27714 28946: gap of unknown length
* 28947 29046: gap of unknown length
* 29047 30187: contig of 1141 bp in length
* 30188 30287: gap of unknown length
* 30288 31532: contig of 1245 bp in length
* 31533 31632: gap of unknown length
* 31633 32859: contig of 1227 bp in length
* 32860 32959: gap of unknown length
* 32960 34118: contig of 1159 bp in length
* 34119 34218: gap of unknown length
* 34219 35277: contig of 1059 bp in length
* 35278 35377: gap of unknown length
* 35378 36613: contig of 1236 bp in length
* 36714 37965: contig of 1252 bp in length
* 37966 38065: gap of unknown length
* 38066 39167: contig of 1102 bp in length
* 39168 39267: gap of unknown length
* 39268 40333: contig of 1066 bp in length
* 40334 40433: gap of unknown length
* 40434 41793: contig of 1360 bp in length
* 41794 41893: gap of unknown length
* 41894 44026: contig of 2133 bp in length
* 44027 44126: gap of unknown length
* 44127 45784: contig of 1658 bp in length
* 45785 45884: gap of unknown length
* 45885 47475: contig of 1591 bp in length
* 47476 47575: gap of unknown length
* 47576 49524: contig of 1949 bp in length
* 49525 49624: gap of unknown length
* 49625 52194: contig of 2570 bp in length
* 52195 52294: gap of unknown length
* 52295 56081: contig of 3787 bp in length
* 56082 56181: gap of unknown length
* 56182 60752: contig of 4571 bp in length
* 60753 60852: gap of unknown length
* 60853 64243: contig of 3391 bp in length
* 64244 64343: gap of unknown length
* 64344 72855: contig of 8512 bp in length
* 72856 72955: gap of unknown length
* 72956 79691: contig of 6736 bp in length
* 79692 79791: gap of unknown length
* 79792 88438: contig of 8647 bp in length
* 88439 88538: gap of unknown length
* 88539 95200: contig of 6662 bp in length
* 95201 95300: gap of unknown length
* 95301 104456: contig of 9156 bp in length
* 104457 104556: gap of unknown length
* 104557 113122: contig of 8566 bp in length
* 113123 113222: gap of unknown length

```

```

* 113223 131437: contig of 18215 bp in length
* 131438 131537: gap of unknown length
* 131538 147555: contig of 16018 bp in length
* 147556 147655: gap of unknown length
* 147656 183113: contig of 35458 bp in length
* 183114 183213: gap of unknown length
* 183214 210273: contig of 27060 bp in length
* 210274 210373: gap of unknown length
* 210374 273658: contig of 63285 bp in length.
FEATURES
source 1. 273658
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="16"
/clone="RP23-10A2"
BASE COUNT 74849 a 54887 c 68561 g 60594 t 14767 others
ORIGIN

```

```

alignment_scores:
Quality: 119.50 Length: 48
Ratio: 3.621 Gaps: 2
Percent Similarity: 68.750 Percent Identity: 43.750
alignment_block:
US-09-316-048-27 x AC079423/rev ..

```

Align seg 1/1 to reverse of: AC079423 from: 1 to: 273658

```

1 ThrAlaCysAlaAlaThrCysThrAlaGlyAlaGlyCysGlyCysCys 17
:::|||||::: ||| ::|||::|
125408 TCTGCTGCGGAGAGAGGCTGC...TGGCGGCTGCTGCGGAGATCGCG 125362

```

```

17 sGlyCysGlyCysThrCysAlaGlyAlaGly..... 27
|||||::: ||| ::|||::|
125361 AGGCTGCGGCTGCTGCTGCGGATCGCTGCGGAGAGCTGCGGG 125312

```

```

28 .....GlyGlyThrCysThrAlaGlyThrGlyCysGlyAlaGly 40
|||||::: ||| ::|||::|
125311 GCTGTGAGGCGGCTGCTGCTGCGGCTCGGATCGGAGTGT 125268

```

seq\_name: gb\_hhg24:LMFLCHR12\_07

seq\_documentation\_block:

WPCOMMENT

Sequence split into 23 fragments LOCUS LMFLCHR12 Accession AL390114

Fragment Name	Begin	End
LMFLCHR12_00	1	110000
LMFLCHR12_01	100001	210000
LMFLCHR12_02	200001	310000
LMFLCHR12_03	300001	410000
LMFLCHR12_04	400001	510000
LMFLCHR12_05	500001	610000
LMFLCHR12_06	600001	710000
LMFLCHR12_07	700001	810000
LMFLCHR12_08	800001	910000
LMFLCHR12_09	900001	1010000
LMFLCHR12_10	1000001	1110000
LMFLCHR12_11	1100001	1210000
LMFLCHR12_12	1200001	1310000
LMFLCHR12_13	1300001	1410000
LMFLCHR12_14	1400001	1510000
LMFLCHR12_15	1500001	1610000
LMFLCHR12_16	1600001	1710000
LMFLCHR12_17	1700001	1810000
LMFLCHR12_18	1800001	1910000
LMFLCHR12_19	1900001	2010000
LMFLCHR12_20	2000001	2110000
LMFLCHR12_21	2100001	2210000
LMFLCHR12_22	2200001	2287483

Continuation (8 of 23) of LMFLCHR12 from base 700001 (AL390114 Leishmania major chrom

```
alignment_scores:      Quality: 119.00      Length: 39
                       Ratio: 4.103      Gaps: 2
                       Percent Similarity: 74.359      Percent Identity: 56.410

alignment_block:
US-09-316-048-27 x LMFICHR12_07/rev ..

Align seg 1/1 to reverse of: LMFICHR12_07 from: 1 to: 110000

      2 AlAcysAlaAlaThrcysThraAlaGlyAlaGlyCysGlyCysCysG1 18
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
34465 GCCTGCTGCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 34416
      18 yCysGlyCysThrcysAlaGlyAlaGlyGlyThrcysThraAlaGlyT 35
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
34415 GTGTGGGTGTGGGTGT...GGGTGTGGGTGTGGG...TGTGGGTGTGGGT 34372
      35 hrcGlyCysGlyAlaGly 40
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
34371 GTGGGTGTGGGTGTGGCA 34355

seq_name: gb_hcg17:AC073826

seq_documentation_block:
LOCUS AC073826 146190 bp DNA HTG 29-JUN-2000
DEFINITION Mus musculus clone RP23-93K16, WORKING DRAFT SEQUENCE, 35 unordered
pieces:
ACCESSION AC073826
VERSION AC073826.1 GI:8810443
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus.
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 146190)
          DOE Joint Genome Institute.
          TITLE Sequencing of Mouse
          JOURNAL Unpublished
          AUTHORS DOE Joint Genome Institute.
          REFERENCE 2 (bases 1 to 146190)
          Direct Submission
          Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
          Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
          -----Genome Center
          Center: Joint Genome Institute
          Center Code: JGI
          Web site: http://www.jgi.doe.gov
          -----
          Project Information
          Center Project Name: 1772840
          Center clone name: RPCI-23_93K16
          -----
          Summary Statistics
          Consensus quality: 118979 bases at least Q40
          Consensus quality: 131063 bases at least Q30
          Consensus quality: 133821 bases at least Q20
          Estimated insert size: 187180; agarose-fp estimation
          Estimated insert size: 142790; sum-of-contigs estimation
          Quality coverage: 8.66 in Q20 bases; agarose-fp estimation
          Quality coverage: 11.35 in Q20 bases; sum-of-contigs estimation.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 35 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *
          * 1 1067: contig of 1067 bp in length
          * 1068 1167: gap of unknown length
          * 1168 2519: contig of 1352 bp in length
```

```
* 2520 2619: gap of unknown length
* 2620 3814: contig of 1195 bp in length
* 3815 3914: gap of unknown length
* 3915 5132: contig of 1218 bp in length
* 5133 5232: gap of unknown length
* 5233 6499: contig of 1267 bp in length
* 6499: gap of unknown length
* 6599: gap of unknown length
* 6600 7609: contig of 1010 bp in length
* 7610 7710: gap of unknown length
* 7710 8792: contig of 1083 bp in length
* 8793 8892: gap of unknown length
* 8893 10241: contig of 1348 bp in length
* 10241 10340: gap of unknown length
* 10341 11532: contig of 1192 bp in length
* 11533 11632: gap of unknown length
* 11633 13022: contig of 1390 bp in length
* 13023 13123: gap of unknown length
* 13123 14177: contig of 1055 bp in length
* 14177 14277: gap of unknown length
* 14178 14277: gap of unknown length
* 14278 15606: contig of 1329 bp in length
* 15607 15706: gap of unknown length
* 15707 16958: contig of 1252 bp in length
* 16959 17058: gap of unknown length
* 17059 18676: contig of 1618 bp in length
* 18677 18776: gap of unknown length
* 18777 20033: contig of 1257 bp in length
* 20034 20134: gap of unknown length
* 20134 21567: contig of 1434 bp in length
* 21568 21667: gap of unknown length
* 21668 23259: contig of 1592 bp in length
* 23260 23359: gap of unknown length
* 23360 25489: contig of 2130 bp in length
* 25490 25589: gap of unknown length
* 25590 27760: contig of 2171 bp in length
* 27761 27860: gap of unknown length
* 27861 31380: contig of 3520 bp in length
* 31381 31480: gap of unknown length
* 31481 33663: contig of 2183 bp in length
* 33664 33763: gap of unknown length
* 33764 35838: contig of 2075 bp in length
* 35839 35938: gap of unknown length
* 35939 38189: contig of 2251 bp in length
* 38189 38289: gap of unknown length
* 38290 41114: contig of 2825 bp in length
* 41115 41214: gap of unknown length
* 41215 44502: contig of 3288 bp in length
* 44503 44602: gap of unknown length
* 44603 48703: contig of 4101 bp in length
* 48704 48803: gap of unknown length
* 48804 55008: contig of 6205 bp in length
* 55009 55108: gap of unknown length
* 55109 60259: contig of 5151 bp in length
* 60260 60359: gap of unknown length
* 60360 67080: contig of 6721 bp in length
* 67081 67180: gap of unknown length
* 67181 76456: contig of 9276 bp in length
* 76457 76556: gap of unknown length
* 76557 86359: contig of 9803 bp in length
* 86360 86459: gap of unknown length
* 86460 98295: contig of 11836 bp in length
* 98296 98395: gap of unknown length
* 98396 111872: contig of 13477 bp in length
* 111873 111972: gap of unknown length
* 111973 122470: contig of 10498 bp in length
* 122471 122570: gap of unknown length
* 122571 146190: contig of 23620 bp in length.
*
Location/Qualifiers
1..146190
  /organism="Mus musculus"
  /db_xref="taxon:10090"
  /clone="RP23-93K16"
  /clone_1fb="RPCI mouse BAC library 23"
BASE COUNT 35412 a 36979 c 36009 g 34377 t 3413 others
```



```

ORIGIN
alignment_scores:
  Quality: 118.00      Length: 39
  Ratio: 4.214         Gaps: 0
  Percent Similarity: 71.795      Percent Identity: 48.718
alignment_block:
  US-09-316-048-27 x AC073826/rev ..
Align seg 1/1 to reverse of: AC073826 from: 1 to: 146190
2 AlAcysAlaAlaThrcysThralaGlyAlaGlyCysGlyCysGlyCysGly 18
  ::::::::::: ||| :::::::::::||||| |||||
139701 ACCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGG 139652
18 yCysGlyCysThrcysAlaGlyAlaGlyGlyGlyThrcysThralaGlyT 35
139651 CTTGGCTCTGGCTGGCTGGCTGGCTGGCGGCGTGGCGCTGGCTGGCT 139602
35 hrcGlyCysGlyAlaGly 40
139601 GTGGCGTGGCGCTGGCG 139585
seq_name: gb_htgl7:AC073748
seq_documentation_block:
LOCUS AC073748 175371 bp DNA HTG 29-JUN-2000
DEFINITION Mus musculus clone RP23-305J2, WORKING DRAFT SEQUENCE, 33 unordered
pieces.
AC073748
AC073748.1 GI:8810365
HTG: HTGS_PHASE1; HTGS_DRAFT.
house mouse.
Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
DOE Joint Genome Institute.
Sequencing of Mouse
Unpublished
2 (bases 1 to 175371)
DOE Joint Genome Institute.
Direct Submission
Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1854210
Center clone name: RPCI-23-305J2
-----
Summary Statistics
Consensus quality: 152503 bases at least Q40
Consensus quality: 164435 bases at least Q30
Consensus quality: 166779 bases at least Q20
Estimated insert size: 166550; agarose-fp estimation
Estimated insert size: 172171; sum-of-ctnigs estimation
Quality coverage: 11.93 in Q20 bases; agarose-fp estimation
Quality coverage: 11.54 in Q20 bases; sum-of-ctnigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 ctnigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the ctnigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1182: ctnig of 1182 bp in length

```

	*	1183	1282:	gap of unknown length
	*	1283	2855:	contig of 1571 bp in length
	*	2854	2953:	gap of unknown length
	*	2954	4183:	contig of 1230 bp in length
	*	4184	4283:	gap of unknown length
	*	4284	5786:	contig of 1503 bp in length
	*	5787	5886:	gap of unknown length
	*	5887	7066:	contig of 1180 bp in length
	*	7067	7166:	gap of unknown length
	*	7167	8238:	contig of 1072 bp in length
	*	8239	9318:	gap of unknown length
	*	9319	9718:	contig of 1380 bp in length
	*	9719	9818:	gap of unknown length
	*	9819	11510:	contig of 1692 bp in length
	*	11511	11610:	gap of unknown length
	*	11611	12959:	contig of 1349 bp in length
	*	12960	13059:	gap of unknown length
	*	13060	14954:	contig of 1895 bp in length
	*	14955	15054:	gap of unknown length
	*	15055	16862:	contig of 1808 bp in length
	*	16863	16962:	gap of unknown length
	*	16963	19187:	contig of 2225 bp in length
	*	19188	19287:	gap of unknown length
	*	19288	21631:	contig of 2344 bp in length
	*	21632	21731:	gap of unknown length
	*	21732	23582:	contig of 1851 bp in length
	*	23583	23682:	gap of unknown length
	*	23683	25787:	contig of 2105 bp in length
	*	25788	25887:	gap of unknown length
	*	25888	27450:	contig of 1563 bp in length
	*	27451	27551:	gap of unknown length
	*	27551	31645:	contig of 4096 bp in length
	*	31647	31745:	gap of unknown length
	*	31747	33311:	contig of 1565 bp in length
	*	33312	33411:	gap of unknown length
	*	33412	34480:	contig of 3069 bp in length
	*	36481	36580:	gap of unknown length
	*	36581	40140:	contig of 3560 bp in length
	*	40141	40240:	gap of unknown length
	*	40241	46716:	contig of 6476 bp in length
	*	46717	46816:	gap of unknown length
	*	46817	52150:	contig of 5334 bp in length
	*	52151	52250:	gap of unknown length
	*	52251	56774:	contig of 4524 bp in length
	*	56775	56874:	gap of unknown length
	*	56875	62585:	contig of 5711 bp in length
	*	62586	62685:	gap of unknown length
	*	62686	72379:	contig of 9694 bp in length
	*	72380	72475:	gap of unknown length
	*	72480	78605:	contig of 6126 bp in length
	*	78606	78705:	gap of unknown length
	*	78706	87900:	contig of 9195 bp in length
	*	87901	88000:	gap of unknown length
	*	88001	98833:	contig of 10832 bp in length
	*	98833	98932:	gap of unknown length
	*	98933	109561:	contig of 10629 bp in length
	*	109562	109661:	gap of unknown length
	*	109662	127036:	contig of 17375 bp in length
	*	127037	127136:	gap of unknown length
	*	127137	139467:	contig of 12331 bp in length
	*	139468	139567:	gap of unknown length
	*	139568	158423:	contig of 18856 bp in length
	*	158424	158523:	gap of unknown length
	*	158524	175371:	contig of 16848 bp in length.
FEATURES	.	Location/Qualifiers		
SOURCE	.	1..175371		
	.	/organism="Mus musculus"		
	.	/db_xref="taxon:10090"		
	.	/clone="RP23-30542"		
BASE COUNT	.	/clone_lib="RPCI mouse BAC library 23"		
ORIGIN	.	44046 a 42634 c 42369 g 43110 t 3212 others		

```

alignment_scores:      Quality: 116.00      Length: 39
                       Ratio: 4.000      Gaps: 2
                       Percent Similarity: 74.359      Percent Identity: 53.846

alignment_block:
US-09-316-048-27 x AC073748 ..

Align seg 1/1 to: AC073748 from: 1 to: 175371

2 AlacysAla1aThrcysThraAlaGlyCysGlyCysGlyCysGly 18
:::|||||:  |||  ::|||:|||||  ||||
33902 ACCTGTGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 32951
18 YCYSGlyCysThrcysAlaGlyAlaGlyGlyThrcysThraAlaGlyT 35
|||||  |||  |||:||||  |||  |||  ::|||
32952 CTGTGCTGTGCTGTG...GGCTGTGCTGTGGC...TGTGCTGTGCT 32995
35 hrcGlyCysGlyAlaGly 40
|||||  |||:||||  |||
32996 GTGGCTGTGCTGTGGC 33012

seq_name: gp_hcg17:AC073748

seq_documentation_block:
LOCUS AC073748 175371 bp DNA HTG 29-JUN-2000
DEFINITION Mus musculus clone RP23-305J2, WORKING DRAFT SEQUENCE, 33 unordered
pieces:
AC073748
VERSION AC073748.1 GI:8810365
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 175371)
DOE Joint Genome Institute.
Sequencing of Mouse
Unpublished
2 (bases 1 to 175371)
DOE Joint Genome Institute.
Direct Submission
Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1654210
Center clone name: RPC1-23_305J2
-----
Summary Statistics
Consensus quality: 152503 bases at least Q40
Consensus quality: 164435 bases at least Q30
Consensus quality: 166779 bases at least Q20
Estimated insert size: 166550; agarose-fp estimation
Estimated insert size: 172171; sum-of-contigs estimation
Quality coverage: 11.93 in Q20 bases; agarose-fp estimation
Quality coverage: 11.54 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1182: contig of 1182 bp in length
* 1183 1282: gap of unknown length
* 1283 2853: contig of 1571 bp in length

```

```

* 2854 2953: gap of unknown length
* 2954 4183: contig of 1230 bp in length
* 4184 4283: gap of unknown length
* 4284 5786: contig of 1503 bp in length
* 5787 5886: gap of unknown length
* 5887 7066: contig of 1180 bp in length
* 7067 7166: gap of unknown length
* 7167 8238: contig of 1072 bp in length
* 8239 8339: gap of unknown length
* 8339 9718: contig of 1380 bp in length
* 9719 9818: gap of unknown length
* 9819 11510: contig of 1692 bp in length
* 11511 11610: gap of unknown length
* 11611 12959: contig of 1349 bp in length
* 12960 13059: gap of unknown length
* 13060 14954: contig of 1895 bp in length
* 14955 15054: gap of unknown length
* 15055 16862: contig of 1808 bp in length
* 16863 16962: gap of unknown length
* 16963 19187: contig of 2225 bp in length
* 19188 19287: gap of unknown length
* 19288 21631: contig of 2344 bp in length
* 21632 21731: gap of unknown length
* 21732 23582: contig of 1851 bp in length
* 23583 23682: gap of unknown length
* 23683 25787: contig of 2105 bp in length
* 25788 25887: gap of unknown length
* 25888 27450: contig of 1563 bp in length
* 27451 27550: gap of unknown length
* 27551 31646: contig of 4096 bp in length
* 31647 31746: gap of unknown length
* 31747 33311: contig of 1565 bp in length
* 33312 33411: gap of unknown length
* 33412 36480: contig of 3069 bp in length
* 36481 36580: gap of unknown length
* 36581 40140: contig of 3560 bp in length
* 40141 40240: gap of unknown length
* 40241 46716: contig of 6476 bp in length
* 46717 46816: gap of unknown length
* 46817 52150: contig of 5334 bp in length
* 52151 52250: gap of unknown length
* 52251 56774: contig of 4524 bp in length
* 56775 56874: gap of unknown length
* 56875 62585: contig of 5711 bp in length
* 62586 62685: gap of unknown length
* 62686 72379: contig of 9694 bp in length
* 72380 72479: gap of unknown length
* 72480 78605: contig of 6126 bp in length
* 78606 78706: gap of unknown length
* 78707 87900: contig of 9195 bp in length
* 87901 88000: gap of unknown length
* 88001 98832: contig of 10832 bp in length
* 98833 98932: gap of unknown length
* 98933 109561: contig of 10629 bp in length
* 109562 127036: contig of 17375 bp in length
* 127037 127136: gap of unknown length
* 127137 139467: contig of 12331 bp in length
* 139468 139567: gap of unknown length
* 139568 158423: contig of 18856 bp in length
* 158424 158523: gap of unknown length
* 158524 175371: contig of 16846 bp in length.
*
FEATURES
     source
         1..175371
             /organism="Mus musculus"
             /db_xref="taxon:10090"
             /clone="RP23-305J2"
             /clone_1fb="RPC1 mouse BAC library 23"
BASE COUNT  44046 a 42634 c 42369 g 43110 t 3212 others
ORIGIN
alignment_scores:

```

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 5 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will

pertussis, bordetella parapertussis, and bordetella bronchiseptica



```

misc_feature      132..141           /note="consensus sequence"
misc_feature      185..190           /note="tcaagt sequence"
promoter          242..248
misc_feature      306..307           /note="TATA-box"
misc_feature      846..847           /note="exon/intron junction"
misc_feature      1043..1045         /note="intron/exon junction"
misc_feature      1159..1161         /note="stop codon"
misc_feature      1312..1314         /note="stop codon"
misc_feature      1347..1353         /note="stop codon"
misc_feature      1461..1466         /note="pot. polyA signal"
misc_feature      1461..1466         /note="pot. polyA signal"
BASE COUNT       453 a      280 c      405 g      453 t
ORIGIN
alignment_scores:
    Quality:   112.50             Length:   29
    Ratio:     5.114              Gaps:     1
Percent Similarity: 75.862        Percent Identity: 65.517

alignment_block:
US-09-316-048-27 x BMHCBRGN ..

Align seg 1/1 to: BMHCBRGN from: 1 to: 1591

10 glyAlaglcysgilycysgcysgilycysgilycysgilycysalagl 26
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1165 GGATGGCGGTTGTGCAGAGTGTGTGCAGAGATGTGCATCTGTTGTA 1214
26 aclyglyglythrCysThrAlaglythrGlyCysgly 38
||||||| ||| :||| | |||||
1215 CGGTGGATGCCGATGCCGATGCCGCTGTGA...GGTGTGTGA 1248

seq_name: gb_v12:OPU39145

seq_documentation_block:
LOCUS      OPU39145      5388 bp      DNA      VRL      18-JUL-1996
DEFINITION Orygia pseudotsugata nuclear polyhedrosis virus DNA polymerase
            gene, complete cds, and ORF 66 gene, partial cds.
ACCESSION  U39145
VERSION    U39145.1 GI:1063687
KEYWORDS   .
SOURCE     Orygia pseudotsugata nuclear polyhedrosis virus.
ORGANISM   Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
            Nucleopolydnavirus.
REFERENCE  1 (bases 1 to 5388)
AUTHORS   Gross,C.H., Wolgamot,G.M., Russell,R.L., Pearson,M.N. and
            Rohrmann,G.F.
TITLE     A 37-kilodalton glycoprotein from a baculovirus of Orygia
            pseudotsugata is localized to cytoplasmic inclusion bodies
JOURNAL    J. Virol. 67 (1), 469-475 (1993)
MEDLINE    93100831
REFERENCE  2 (bases 1 to 5388)
AUTHORS   Ahrens,C.H., Carlson,C. and Rohrmann,G.F.
TITLE     Identification, sequence, and transcriptional analysis of lef-3, a
            gene essential for Orygia pseudotsugata baculovirus DNA replication
JOURNAL    Virology 210 (2), 372-382 (1995)
MEDLINE    95343549
REFERENCE  3 (bases 1 to 5388)
AUTHORS   Ahrens,C.H. and Rohrmann,G.F.
TITLE     The DNA polymerase and helicase genes of a baculovirus of Orygia
            pseudotsugata
JOURNAL    J. Gen. Virol. 77 (Pt 5), 825-837 (1996)
```

[illegible]

alignment\_scores:                   Quality: 112.50                   Length: 48  
                                       Ratio: 3.879                   Gaps: 3  
 Percent Similarity: 60.417       Percent Identity: 45.833

alignment\_block:  
 US-09-316-048-27 x OP039145/rev ..

Align seg 1/1 to reverse of: OP039145 from: 1 to: 5388

4   AlaAlaThrCysThrAlaGly.....AlaGlyCysGlyGlyCysCys.. 17  
       |||||:||||       |||||       |||||       |||||       |||||  
 3419 AGCGCCAACTGTTGCGCGCGCGCGCGCTGCGCGCGCTGCGCGCGCTGCGG 3370  
       |||||:|||||       |||||       |||||       |||||       |||||  
 18   .....GlyCysGlyCysThrCysAlaGlyAlaGlyGlyGlyThrc 31  
       |||||:|||||       |||||       |||||       |||||       |||||  
 3369 CGGCTCGCGCGCGCTGCGCGCTGCTGCGCGCGCTGCGCGCGCTGCTGCT 3320  
       |||||:|||||       |||||       |||||       |||||       |||||  
 31   ys.....ThrAlaGlyThrGlyCysGly 38  
       |||||:|||||       |||||       |||||       |||||       |||||  
 3319 GCGCGCGCTGCGCGCTGTTGGTTCGCGAGCTGCTGCGCGC 3276

seq\_name: gb\_v12:OP075930

seq\_documentation\_block:  
 LOCUS   OP075930   131993 bp       DNA       circular   VRL       06-MAR-1998  
 DEFINITION   Orygia pseudotsugata nuclear polyhedrosis virus complete genome.  
 ACCESSION   U75930  
 VERSION   U75930.1   GI:2934903  
 KEYWORDS  
 SOURCE  
 ORGANISM   Orygia pseudotsugata nuclear polyhedrosis virus.  
             Orygia pseudotsugata nuclear polyhedrosis virus  
             Viruses; dsDNA viruses, no RNA stage; Baculoviridae;  
             Nucleopolyhedrovirus.  
 1 (bases 1 to 131993)  
 Ahrens,C.H., Russell,R.L., Funk,C.J., Evans,J.T., Harwood,S.H. and  
 Rohrmann,G.F.  
 The sequence of the Orygia pseudotsugata multinnucleocapsid nuclear  
 polyhedrosis virus genome  
 Virology 229 (2), 381-399 (1997)  
 2 (bases 1 to 131993)  
 Rohrmann,G.F.  
 Direct Submission  
 Submitted (23-OCT-1996) Oregon State University, Agricultural  
 Chemistry, Corvallis, OR 97331-7301, USA  
 3 (bases 1 to 131993)  
 Rohrmann,G.F.  
 Direct Submission  
 Submitted (06-MAR-1998) Oregon State University, Agricultural  
 Chemistry, Corvallis, OR 97331-7301, USA  
 Sequence update by submitter  
 On Mar 6, 1998 this sequence version replaced gi:1911246.  
 similar to Autographa californica nuclear polyhedrosis virus  
 (ACMNPV) complete genome: GenBank Accession Number L22858.  
 FEATURES  
 source  
 1. 131993  
   /organism="Orygia pseudotsugata nuclear polyhedrosis  
   virus"  
   /db\_xref="taxon:10450"  
   /complement(123..947)  
   /note="ORF1: p1, ACMNPV ORF10 homolog"  
   /codon\_start=1  
   /product="protein kinase 1"  
   /protein\_id="AAC59000.1"  
   /db\_xref="GI:1911247"  
   /translation="MDAALOSLSRFVADCAVLAKPVYNGRGKMDVLHHRPTSKLFL  
   RKTIAHSFSDELINVDLMSDPFVDMFCYSSPTAMALVMDYVPCPDILFTLQTO  
   GALDNLAVNVIRQLDNLNDLHNATGYIHNDAVLENVLYFGARDRYVLCIDYGLCKRE  
   HSPVHGTELEFSPKIRHNHARSFDMYAVGVALYKLTGGRHPPERSADVEDLAS  
   MRROQYNDPAALKNVNLMARDFVCLTRNFECRSTDYKQIAKHSFLASRDHYI"

CDS  
 946..2370  
   /note="ORF2: ACMNPV ORF9 homolog"  
   /codon\_start=1  
   /product="1629-capsid"  
   /protein\_id="AAC59001.1"  
   /db\_xref="GI:2934904"  
   /translation="MEROYOSVRSYLLINNOHNAIAGPFLQRYVAGPEAHSGRNVGD  
   RAVRLNQAIVDLIKLAEDYADPTAYVADQPEASSRHEATLNMRLLILGVDPDAR  
   RNENSVLARIETALLRVYVNDAEVNVLSGDFEYESKISTYQGFACQPTPTASAOQTO  
   TSLRPQTSLEPRTQASLPQPTFPDQPEVMSPPSVHTTPAILQTTQPPATVDFSRP  
   SDEFVYVPGKRAVADTRKPPVPEPHLSRSSVATNAAGATPVAAPPPVADTSMP  
   VTSMPPPPPPPPADVTTSMPPPPVLDLSTMPPPPPVADLSTMPPPVADLSTMP  
   INNLIDMAVETKNNAGNRSALDDQKCKTKTKQPPADGAPATDRSTYLSLEIRQ  
   GKTIKLRKIEDOSTOTLLKVDVDTKTKILKNFVYNIDRISKQDEEKDRDITIT  
   KRRAVEHTDQNSGNNSDDPRD"  
   /complement(2333..3270)  
   /note="ORF3: p1, ACMNPV ORF8 homolog"  
   /codon\_start=1  
   /product="polyhedrin"  
   /protein\_id="AAC59002.1"  
   /db\_xref="GI:1911249"  
   /translation="MPTSYRPTGRTTYVYVNTKYNKNGSYKAKKRRKKHLEHDEE  
   KHLDPLHYVAEDPFLGPGKNOKLTLEKELIRNVKPDIMKLDIVMSGKEFLRETWR  
   VEDSEPIYVNDQVMDVFLVNMRRPTPRNRCYKFLAOLHDCDYPVHEVYRIVEPSY  
   VGMNNEYRISLAKKGGCCPIMNIIHAETYNSEFVNRVIMENFYKPIYVIGTDSSEE  
   ELIEVSLVFKKEFAPDAPLFTGPAY"  
   /complement(3350..3541)  
   /note="ORF4"  
   /codon\_start=1  
   /product="unknown"  
   /protein\_id="AAC59003.1"  
   /db\_xref="GI:1911250"  
   /translation="MRRKNNGASQYRPHHTTVYHKKHRLQSLCVMCDQDQSMPSHAKGL  
   MFEHMFCLRLIYVQKIKH"  
   /complement(3698..4093)  
   /note="ORF5"  
   /codon\_start=1  
   /product="unknown"  
   /protein\_id="AAC59004.1"  
   /db\_xref="GI:1911251"  
   /translation="MGASQNSEQKILKLFYRWSSQTGAALDEKDLHCLYDLERFVGA  
   HLNKRADDKARKKCAERKALRVEIADRIIMLEAANAAPACADGGRWSTLSRAOLDI  
   AREKEIVDIRIQLKODSLKODRLKKR"  
   /complement(4125..4739)  
   /note="ORF6: 1ef-2, ACMNPV ORF6 homolog"  
   /codon\_start=1  
   /product="late expression factor 2"  
   /protein\_id="AAC59005.1"  
   /db\_xref="GI:1911252"  
   /translation="MERVWNPAGIDGLKRSSEYLVDPHDFVGLVTLSPYVFERGLE  
   VRMSGRLIALAPKQEPQPAVRPRPQRRRVCLAKADGASQSLAKVITAEVNASMP  
   PCMKTYADLSSARGNKKRFRFNCYLAIVITCTCKTACILGALLHFRIMDAKVY  
   GEVTHLIKADVYKPSNCAKMKRYTKLCPASCKGLNPICNF"  
   /complement(4742..4975)  
   /note="ORF7: ACMNPV ORF5 homolog"  
   /codon\_start=1  
   /product="unknown"  
   /protein\_id="AAC59006.1"  
   /db\_xref="GI:1911253"  
   /translation="MNRPTMNTAAVTTDYDQELRRLNSLRSVHLELCTRSATGED  
   CNRFLGADRAPAVIVKAAANGCOHSSLICDV"  
   /5020..5460  
   /note="ORF8: ACMNPV ORF4 homolog"  
   /codon\_start=1  
   /product="unknown"  
   /protein\_id="AAC59007.1"  
   /db\_xref="GI:1911254"  
   /translation="MPGARVYRFSRLRLTQEFKENVAVHVDHLMGLRALIDGKYSADV  
   RRGFLSRNVLVSCAMVNVQVYADPTIMRQDPTIYFRVQCNCHMADVPAPADDS  
   IARVLECGAVLVADHPDVFEGTEEGVNLLEVQRI MAGDL"  
   /complement(5669..6151)  
   /note="ORF9: p1p-2"  
   /codon\_start=1

CDS  
 946..2370  
   /note="ORF2: ACMNPV ORF9 homolog"  
   /codon\_start=1  
   /product="1629-capsid"  
   /protein\_id="AAC59001.1"  
   /db\_xref="GI:2934904"  
   /translation="MEROYOSVRSYLLINNOHNAIAGPFLQRYVAGPEAHSGRNVGD  
   RAVRLNQAIVDLIKLAEDYADPTAYVADQPEASSRHEATLNMRLLILGVDPDAR  
   RNENSVLARIETALLRVYVNDAEVNVLSGDFEYESKISTYQGFACQPTPTASAOQTO  
   TSLRPQTSLEPRTQASLPQPTFPDQPEVMSPPSVHTTPAILQTTQPPATVDFSRP  
   SDEFVYVPGKRAVADTRKPPVPEPHLSRSSVATNAAGATPVAAPPPVADTSMP  
   VTSMPPPPPPPPADVTTSMPPPPVLDLSTMPPPPPVADLSTMPPPVADLSTMP  
   INNLIDMAVETKNNAGNRSALDDQKCKTKTKQPPADGAPATDRSTYLSLEIRQ  
   GKTIKLRKIEDOSTOTLLKVDVDTKTKILKNFVYNIDRISKQDEEKDRDITIT  
   KRRAVEHTDQNSGNNSDDPRD"  
   /complement(2333..3270)  
   /note="ORF3: p1, ACMNPV ORF8 homolog"  
   /codon\_start=1  
   /product="polyhedrin"  
   /protein\_id="AAC59002.1"  
   /db\_xref="GI:1911249"  
   /translation="MPTSYRPTGRTTYVYVNTKYNKNGSYKAKKRRKKHLEHDEE  
   KHLDPLHYVAEDPFLGPGKNOKLTLEKELIRNVKPDIMKLDIVMSGKEFLRETWR  
   VEDSEPIYVNDQVMDVFLVNMRRPTPRNRCYKFLAOLHDCDYPVHEVYRIVEPSY  
   VGMNNEYRISLAKKGGCCPIMNIIHAETYNSEFVNRVIMENFYKPIYVIGTDSSEE  
   ELIEVSLVFKKEFAPDAPLFTGPAY"  
   /complement(3350..3541)  
   /note="ORF4"  
   /codon\_start=1  
   /product="unknown"  
   /protein\_id="AAC59003.1"  
   /db\_xref="GI:1911250"  
   /translation="MRRKNNGASQYRPHHTTVYHKKHRLQSLCVMCDQDQSMPSHAKGL  
   MFEHMFCLRLIYVQKIKH"  
   /complement(3698..4093)  
   /note="ORF5"  
   /codon\_start=1  
   /product="unknown"  
   /protein\_id="AAC59004.1"  
   /db\_xref="GI:1911251"  
   /translation="MGASQNSEQKILKLFYRWSSQTGAALDEKDLHCLYDLERFVGA  
   HLNKRADDKARKKCAERKALRVEIADRIIMLEAANAAPACADGGRWSTLSRAOLDI  
   AREKEIVDIRIQLKODSLKODRLKKR"  
   /complement(4125..4739)  
   /note="ORF6: 1ef-2, ACMNPV ORF6 homolog"  
   /codon\_start=1  
   /product="late expression factor 2"  
   /protein\_id="AAC59005.1"  
   /db\_xref="GI:1911252"  
   /translation="MERVWNPAGIDGLKRSSEYLVDPHDFVGLVTLSPYVFERGLE  
   VRMSGRLIALAPKQEPQPAVRPRPQRRRVCLAKADGASQSLAKVITAEVNASMP  
   PCMKTYADLSSARGNKKRFRFNCYLAIVITCTCKTACILGALLHFRIMDAKVY  
   GEVTHLIKADVYKPSNCAKMKRYTKLCPASCKGLNPICNF"  
   /complement(4742..4975)  
   /note="ORF7: ACMNPV ORF5 homolog"  
   /codon\_start=1  
   /product="unknown"  
   /protein\_id="AAC59006.1"  
   /db\_xref="GI:1911253"  
   /translation="MNRPTMNTAAVTTDYDQELRRLNSLRSVHLELCTRSATGED  
   CNRFLGADRAPAVIVKAAANGCOHSSLICDV"  
   /5020..5460  
   /note="ORF8: ACMNPV ORF4 homolog"  
   /codon\_start=1  
   /product="unknown"  
   /protein\_id="AAC59007.1"  
   /db\_xref="GI:1911254"  
   /translation="MPGARVYRFSRLRLTQEFKENVAVHVDHLMGLRALIDGKYSADV  
   RRGFLSRNVLVSCAMVNVQVYADPTIMRQDPTIYFRVQCNCHMADVPAPADDS  
   IARVLECGAVLVADHPDVFEGTEEGVNLLEVQRI MAGDL"  
   /complement(5669..6151)  
   /note="ORF9: p1p-2"  
   /codon\_start=1

```
/product="protein tyrosine phosphatase 2"
/protein_id="AAC59008.1"
/db_xref="GI:1911255"
/translation="MYEASQIDEHVEGVGGYGDNEAMLRFEINHAIAVYILSDSDVA
PIRHALGPGVDHIIHVECEAPDCEALFNPMALDYAVRRIGEGKRIYVHYACASR
SAALAVYILMSRQMPYODALNAVOSKRRAVAINDFVRLATRCSTRVYNDELKFOIV
"
complement(6129..6791)
/feature="ORF10: ptp-1, AcMNPV ORF1 homolog"
/codon_start=1
/product="protein tyrosine phosphatase 1"
/protein_id="AAC59009.1"
/db_xref="GI:1911256"
/translation="MLTSRPMAYIYRTSERNGCGAMCTNGVRRRQVHLPSVARAM
FPDRHETACAVIEGTSLCEKEVLALEFYTSDEDRWTASVLAHSLALAVI
DLTNTARYDGQAWMKGLIYKIRVPGRAVDVDDIVAEFTVDEFPFRCPCTMVAV
HMTGLNRSGLYICRYMVERLGSVPTDAIARETARKGHIKIENTYILQDLARKHVRGO
FN"
6819..7814
/feature="ORF11: AcMNPV ORF11 homolog"
/codon_start=1
/product="unknown"
/protein_id="AAC59010.1"
/db_xref="GI:1911257"
/translation="MSLSKLVAVAYGSYNLPDRYGESEYLYRIVHEHNTYAVSN
ASCVRDITATACNLGSHLCEVDARQLDVSFAVRLISMFEGCDATGACADMORALA
DIDRHAPLARARRGRANIEFALDADIPSDVYNAALIGTIGRMHPRGSGLARAVDF
DPDIRADGMVHMKFCVLTVMHLYACGAVPAGCATLRVAVAKHIGPNDEGCNAPALAA
VYGRFCATGREHFAHKTACHMILTFQFMNDLTPADERPCGVTKDEGRCKDYTD
LRTHADALYIHGTDRKNALFDLLCVNASIDIDCDYCVNKEFYAQNKKYKM"
complement(8180..9142)
/feature="ORF12: AcMNPV ORF13 homolog"
/codon_start=1
/product="unknown"
/protein_id="AAC59011.1"
/db_xref="GI:1911258"
/translation="MESWMEGMMNADEQVNAEPDEQAVRRYAVDQRAHSLVRRDVE
RCHPFVFKRYVLDETAGCCSVDECGKGLISHDLORCNFDRQHVQLINELVYGAP
PAQPDLSIGTATKGLIOLLOQSPFANKNEVLLVAGNCKDHDNDLDTKTAVALNHV
KTLNTNSDKFTSAHKSFKLEVAGREFQEPQRLQDTLKNALQCAAPRTATGAVFPR
DVTKHPIHAEVMEGRVEDRGVQIARFGOEHEFRKRKLEFEEGMDVDVGRAPNPLLA
VHCIEEPVANGHGIIRLPLKVIYEVCAVNAHDIIVKAILNKT"
complement(9094..9825)
/feature="ORF13: left-1, AcMNPV ORF14 homolog"
/codon_start=1
/product="late expression factor 1"
/protein_id="AAC59012.1"
/db_xref="GI:1911259"
/translation="MAPCKYTPPERVOMMMADAIVNDSRLAFMTDRPMVHAHNFEDS
AAQLFAYIVKNSIDSVHVKPLEEGGREWYIDAFKDCADKAEMLKVNVGATAPMLFF
EKEEDAVORIMSGNKGPHLMKFCGKPKMDAPKSLREHMFVVPKOPAKLYSGDILRG
SFADCVRAVHMTIGDARBVLRYMPVDVDRVPCNANKQIRAPFSYNKGGDYSKCL
TQOIOORIKACAGCIAGGTPPTSK"
9893..11362
/feature="ORF14: egt, AcMNPV ORF15 homolog"
/codon_start=1
/product="ecdysteroid UDP-glucosyltransferase"
/protein_id="AAC59013.1"
/db_xref="GI:1911260"
/translation="MVFLIALTLATGARAASIIAVLPTPAYSHHVYRAYVHALVK
NCHNTVIKPOLLDYADDECGRVEQIDADMSAOQYKLVASSGVFRKRGVADETTV
TADNYMGLIEMFKDOFDNANVRFLSTNRTFDAYVVEAFADYALVFGHLEFRAPVTOI
AGVGLAEFEERRAVARHPLHYPTEGAALTRRGALSEMRLLNEPELLARSDLL
KOQFKSTPTIROLNDNOVOLLNLNHPYDNNRPVPSVQYLGGLHLAQLPORDA
PLERLNEVDGAVYVFSGSDTNSIAEFLQMLDTEFANLNNTVILMKVDDAAVAS
VALPRNVLAQKMSQTAFLRHNVAFAVYQAGIOSDEALQARVPWVCLPMMDQFHH
ARKLQFGVARRALDTPAAVAPOLQAIREVIADEGAYRAIDKLRRAVVEHDAAPDEKA
VKETERVTKENNDDVMNPARSLKTTAANMAYSDYFVRFPL"
11547..12140
/feature="ORF15: AcMNPV ORF16 homolog"
/codon_start=1
/product="unknown"
/protein_id="AAC59014.1"
```

```
/db_xref="GI:1911261"
/translation="METAPPIISVAPPKGAVCAVVRVVTTVSDSGNNEDRLQ
IVAQLQRTLNFSKLSQLORRVRNMOKLIRKNSVINIAARLTQKTKHFAVYIR
KNVHTTSGSEQFVORVLELCANGCGEOVFCARRDCARDRRRAVALATALGACVVA
SNAKRFEIIDEKILVSAKLIVQVYLHDSHSDTCAD"
12109..12732
/feature="ORF16: AcMNPV ORF17 homolog"
/codon_start=1

alignment_scores:
  Quality: 112.50      Length: 48
  Ratio: 3.879        Gaps: 3
  Percent Similarity: 60.417  Percent Identity: 45.833

alignment_block:
  US-09-316-048-27 x OPV75930/rev ..

Align seg 1/1 to reverse of: OPV75930 from: 1 to: 131993

4 AlaAlaThrcCysThrAlaGly.....AlaGlyCysGlyCysCys.. 17
:::||||:|||||:::|||||
58315 AGCCCACTGTTTCGGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGG 58266

18 .....GlyCysGlyCysThrCysAlaGlyAlaGlyGlyGlyThrc 31
|||||
58265 CGGCTGCGGCGGCTGCGGCGGCTGCTGCGGCGGCTGCGGCGGCTG 58216

31 yS.....ThralaGlyThrGlyCysGly 38
|||
58215 GCGGCGGCTGCGGCTGTTGGGTTGTCGGAGAGGCTGCGGCGG 58172
```







```

358 lysAlaIThrValThrAlaMetIleAlaArgLLeuLeuLeuTyrGlyGlyTh 374
774 ..... 774
374 rSerProThAlaGlu...ThrIleLeuLysSerAsnIleSerSerGlyH 390
||||| :::::||||| ::::: |||
775 .TCTCCTGGTGAAGAAGTACACTACTGAGAAAGAAGCCTCCTTGGT... 819
390 iValAlaProHisGlyProArgThhArgProSerGluGlnLeuTyrTyrLeu 406
||||| ::::: |||
820 .....ATTCCAAATACAGATTACATCCAGCTGCTAGTGAATTT 858
407 SerArgAlaGlnIlePheGlnValGluTyrLysAspPheProLys..... 421
::::: ||||||| ::::: ||| :::::
859 GCCAAGCAACAAAGCTTTTAATATACATATTTGGATATATAGATGAAGCTGAG 908
422 ...AsnAsnLysAsnGluCysValSerLeuIleAsnCysSerSerGlnP 437
||||| ::::: ||| ::::: ||| :::::
909 CGGCAGTGGCAACTATTCATGT.....CTTGCAGAACTGTGCCAAGCC 952
437 rObProLeuValSerHisGlyIleGlyLysAspValGluSerCysHisAsp 453
||| ||| ||| ||| :::::
953 CCATACACAGCTGTGCATGCGCTCCGGTATCTCTGTGGCAATGACAAAGT 1002
454 MetAlaIleLeuAsnIleLeuLysLeu..... 463
||||| ||| ||| :::::
1003 GATGCAGAGCTCACATGCTTTCAGATTAATAAGATTAATAGCAGAAAGAAA 1052
464 .....SerGluLeuAspGlnGlnSerThgluMetProArgThrg 477
||||| ::::: ||| :::::
1053 GTAATCTGGAGAGCACTTAATAAATCTTTCAGTGCACATTAATAAAGTTCC 1102
477 LysAsnGlyPro 480
::: ||| |||
1103 CCTCTGGCCCC 1113

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-933-750C-95

seq_documentation_block:
; Sequence 95, Application US/08933750C
; Patent No. 5932442
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750C
; FILING DATE: September 23, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.

```



1131 ACTTAAAAATCTTTCAGTAGACATATAAAAGTCCCTCTGCCCC 1177

seq\_name: /cgn2\_6/ptodata/2/1na/6\_COMB.seq:us-09-234-613-95

seq\_documentation\_block:

Sequence 95, Application US/09234613

Patent No. 6132973

GENERAL INFORMATION:

APPLICANT: Lal, Preeti

APPLICANT: Hillman, Jennifer L.

APPLICANT: Bandman, Olga

APPLICANT: Shah, Purvi

APPLICANT: Au-Young, Janice

APPLICANT: Yee, Henry

APPLICANT: Guegler, Karl J.

APPLICANT: Corley, Neil C.

TITLE OF INVENTION: HUMAN REGULATORY MOLECULES

NUMBER OF SEQUENCES: 98

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/234,613

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/933,750

FILING DATE: September 23, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: Pf-0356 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

TELEX:

INFORMATION FOR SEQ ID NO: 95:

SEQUENCE CHARACTERISTICS:

LENGTH: 1904 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: THPLAZS08

CLONE: 2754573

US-09-234-613-95

alignment\_scores:

Quality: 212.00

Ratio: 1.044

Percent Similarity: 40.681

Percent Identity: 22.244

Length: 499

Gaps: 16

Alignment block:

US-09-316-048-11 x US-09-234-613-95 ..

Align seg 1/1 to: US-09-234-613-95 from: 1 to: 1904

46 GYGLVGLnclnphèaànglyLysgLyLysmetArgProValLysH 62

17 GGTGGGGGGCTGGANNCTCTCCGGGAGCACGACCCGGGAGCATGCA 66

62 SASPALAProAlaArgAlaLeuArgThrLeuGlnSerGluProLeuPro. 78

67 CGTGTGGACACGACTTCACGTGAGCTTGGACAGAACTAATCTCCCTC 116

78 ..... 78

117 CATCTCCAACTTCTTGACAAAGCTCGCTCTTCTCCAGCCGTGAGCCG 166

79 ..... 79

167 TCCCTTCTGCGCATGTGCCAGACGACGACCGCGGAGCCCGCGCT 216

86 GGIUALAGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUG 93

217 GGAGCGCGGAGACGTGGACCTTCAGTTGGGGAAGATGATACAGCTA 266

94 ..... 94

267 AGCCAGGAAACACCGATTTCAGCTATTACGCAATACGCAATGAAAGC 316

108 ArgAnLeuProValAsnPhelValAlaArgLysSerGlyProProH 124

317 AAGACATCCCACTT...TATGAATGTGAAGATGTGATGTGCAATPAC 363

124 SMeLysAsnPhelValThrArgValSerValGlyGluPheValGlyG 141

364 CGTGGCCACTTTCACCTTCAGACTACCTTGGTGAATTAACCTGCACAG 413

141 LYGILUGLysSerLysLysLysSerLysLysLysLysLysLysLys 157

414 GTGAGGTACAGTAAGAGCTGGCAACATGAGCTGAGAGGCTGCC 463

158 LeuGILGlnLeuArgArgLeuProProLeuProAlaValGluArgVal 174

464 ATA..... 466

174 SPoATgLIeLysLysSerGlnProHrCysLysThrAlaProasp. 190

467 ....AACATTGTGAAGCCAAATGCAATGATTTCTTGTGCTCCGACC 512

191 ..... 198

513 CCTTAATGCCGTGACCCCTCCACCAACCAAGACAGCTTAATCTTAAT 562

199 SerArgLeuAlaGlnIleGlnGlnAlaLysLysGlnLysGluProGly 215

563 GGTTCATTAACAGGAATTTGGCTATTCAATGCTGGAGACTTCTGAATA 612

215 rMeLLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMet 232

613 TACCCTTTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 662

232 InValLysValGlyHisHisThrAlaGluGlyValGlyThrAsnLys 248

663 TTTGCAAGGCTAGAGTCATTTATGAAACCTGGAAGGGGAGATCAAAAAG 712

249 ValAlaLysArgAsnAlaAlaGlnAsnMetLeuGluIleLeuGlyPhe 265

713 CAGCCAAAGAAATGCTGCTGAGAAATTTCTT..... 745

265 sValProGlnAlaGlnProAlaLysProAlaLeuLysSerGluGluLys 282

745 ..... 745

282 hrProValLysLysProGlyAspGlyArgLysValThrPhePheGluPro 298

746 ..... 763

299 SerProGlyAspGlnLysGlnLysThrSerAsnLysAspGluGluPheArg 315

764 TCTTCA.....GAGAACCAATTTCT..... 784

315 tProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuProMetValP 332

784 ..... 784

```

3332  roglvalalaglnalvalglvalsercngly.....hshis 345
      ::::::::::::::||| |||
785   ....TTAACAAATGATGAGCATCTTTAGAGTACTGGCATCC 829
      :::::::::::::::
346  ThrLysAspPheThrArgAlaIalProAsnProAlaLysAlaThrValTh 362
      :::::
830  TWAGGAAT.....TCCTGGTGAA 838
      ::::::::::
362  rAlaMetIleAlaArgIleuLeuTyrGlyThrSerProThrIaG 379
      ::::::::::
839  ....ThrlleuLysSerAsnIleSerSerGlyHisValProHisGly 851
      :::::::::::::::
379  lu...ThrlleuLysSerAsnIleSerSerGlyHisValProHisGly 394
      :::::::::::::::
852  AGATCACTTACTGAAAAGACGCTCTTAGT.....ATT 886
395  ProArgThrArgProSerGluGluLeuTyrTyrLeuSerArgAlaGlnI 411
      :::::::::::::::
887  CCAAAATACGATTTACATCCAGCTCTTAGGAATTCCAAGAACAG 936
411  yheGlnValGluTyrLysAspPheProLys.....AsnAsnLysA 425
      :::::::::::::::
937  TTTTAATATACATATTGGATTTGAATGAATGACGCGCAATGCAAT 986
425  sngLysValSerLeuIleAsnCySerSerGlnProLeuValSer 441
      ::::::::::
987  ATCAATGT.....CTTGCACTGTGCCACAGCCCATCAGACTGTGT 1030
442  HisGlyIleGlyLysAspValGluSerCyHisAspMetAlaIleuAs 458
      :::::::::::::::
1031  CATGGCTCGCGTATCTCTGTGGCAATGCACAAAGTGATGCAGTCAAA 1080
458  nIleuLysLeuLys.....Serg 465
      ::::::::::
1081  TGCTTTCACGATTTAAAGATAATAGCAGAAAGATAATCTGGAGCA 1130
465  luleuAspGlnIleSerThrGluMetProArgThrGlyAsnGlyPro 480
      ::::::::::
1131  ACTTAAATAATCTTCAGTAGACACATTAAGAAGTCCCTCGGGCCC 1177
seq_name: /cgn2_6/ptodate/2/ina/5A.COMB.seq:US-08-280-443-1
seq_documentation_block:
Sequence 1, Application US/08280443
Patent No. 5643778
GENERAL INFORMATION:
APPLICANT: Nishikura, Kazuko
TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr, P.O. Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/280,443
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/197,794
FILING DATE: 17-FEB-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215

```

```

? REFERENCE/DOCKET NUMBER: WSP49AUSA
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 215-540-9206
? TELEFAX: 215-540-5818
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 6671 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: unknown
? MOLECULE TYPE: cDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 155..3832
?
? US-08-280-443-1

alignment_scores:
      Quality: 171.00      Length: 526
      Ratio: 0.710      Gaps: 23
      Percent Similarity: 45.817      Percent Identity: 21.483

alignment_block:
US-09-316-048-11 x US-08-280-443-1 ..

Align seg 1/1 to: US-08-280-443-1 from: 1 to: 6671

55 LysMetArgProValLysHisAspAlaProAlaArgala..... 68
   ::::::::::::::::::::
1445 AGACGTGAACACCCTGTCATTACAAATGAGCCCTCAAAAGCAGGATATGT 1494
   69 .....LeuA 70

1495 TGACCTTTGAATAATGGCCAGTGGCCACAGATGACATCCACAGATCCTTGA 1544
   70 rghrLeuGlnSerGluProLeuProGluArg.....LeuGluValAsn 84
   ::::::::::::::::::::
1545 ATTAGTATCCCGCGACACACAGGTGAGTTTGAGCCATCATGGAATGCC 1594
   85 Gly.....ArgGluAlaG1 89
   ::::::::::::::::::::
1595 TTCCTTGACAGTCATGGCTTGGCCACGGTGTTCACCTTACAAAGAACTGAC 1644
   89 uGluGluAsnLeuAsnLysSerGluLieserGluValPheGluIleAla 106
   ::::::::::::::::::::
1645 AGAGTCCAGACAGCTG...AAGAACCCTCATACAGCGGCTGTAAATAATGCC 1691
   106 euLysArGAsnLeuProValAsnPheGluValAlaArgGluSerG1Pro 122
1692 AGTTCGCTAGTCAACCTGTGAGTTCAACATGATAGACAGAGTGGACCA 1741
   123 ProHisMetLysAsnPheValThrArgValSerVal...GlyGluPheVa 138
   ::::::::::::::::::::
1742 CCCCATGAACCTGGATTAAATTCACAGGTGTGCTCATATGGCCAGAGTT 1791
   138 lGluGluGlyGluGlyLysSerLysLysLieserLysLysAsnAlaAla 154
   ::::::::::::::::::::
1792 TCCCCAGCTGAGAGCTGGAAGAAAGATGGCCAAACGAGATCAGCTTA 1841
   155 .....ArgAlaValLeuGlnGluInLeuArGluProProLeu 167
1842 TGAAGCCATGACAAATTCCTGTAGAGGAACCCAAAGCCAAAGCAGACATGGA 1891
   168 ProAlaLeuGluArgValLysProArgLleLysLysSerGlnProTh 184
   ::::::::::::::::::::
1892 AATATCGAAGAAATCATCCCACTATTCCACAGAGAAAGATATGAGAAAGAC 1941
   184 rCysLys.....ThrAlaProAspTyrLleGlnGly 195
   ::::::::::::::::::::
1942 TGGAGAGTCCCAAGACCCCAACCCCTTGAGCAACATCTTTCTTTGCGGA 1991
   195 eTAsnProIleSerArGLeuAlaGlnIleGlnGlnAlaLysLysGluLys 211
   ::::::::::::::::::::

```

```

1992 AGAGCCCCGTCACACACCTGCTGATGATGAC...AAATTGGGGAAC 2038
212 GluProGluTyrMetLeuLeuThrGluArgGlyLeuProArgArgGly 228
2039 TCCTGGGAATTCGCTCTCGTCCAAAGAGGCCGTCGCCATGACCCAA 2088
228 upheValMetGluValValGlyValHisThrAlaGluGlyValGlyTyr 245
2089 GTTCCAAATCTGCTGTCAGTGGAGGCCCAACTTCCCAAGTGTGAGTG 2138
245 hr...AsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGlu 260
2139 CTCCCAAGCAAGAAAGTGGCAAGACAGATGCCGACAGAACCCATGAAG 2188
261 IleLeuGlyPheLysVal.....ProGly 268
2189 GCCCTCATGGGGAGGCGACCAACTCCATGGCTTGTGATAACGAGCTGA 2238
268 nAlaGluProAlaLysProAlaLeuLysSerGluGlyTyrProValL 285
2239 AGGTATGATCTCAGACGACCTGATGATGATGATGATGATGATGATGAT 2288
285 yslLysProGlyAspGlyArgLysValThrPhePheGluProSerPro... 300
2289 AGGTCAAGAAATGGCGAGCTGTGATACCTGAACCAACCTCTGTG 2338
300 ..... 300
2339 GGTGGCCTTTTGGATACGCCGCTCCATGGCTTGTGCTGTAATTCAA 2388
301 ...GlyAspGluAsnGlyThrSerAsnLysAspGluGluPheArgMetP 316
2389 GTTGTGCAGCACTCCGAGCCTCTCAGCAG.....C 2420
316 roTyrLeuSerHisGlnGlnLeuProAlaGly..IleLeuProMetVal 331
2421 CCAGATGCTGTTTACCAAGCAAAAGTTGGGGGTGCTGTGTTCCAGCCGTC 2470
332 .....ProGluValAlaGlnAla..... 337
2471 TGGGCACACAGCAAGACAGCAAGCAAGCAAGCAAGCAAGCAAGCTCT 2520
338 .....ValGlyValSerGlnGlnHisThrLysAspPheThr 351
2521 CGGTGCTGTGATGGGAGAGCAAGCAAGCAAGCAAGCAAGCAAGCTTCA 2570
351 rgaAlaAlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArg 367
2571 AGTTAAACCCA.....GTGACAGGGCCACGCTCAGAGA 2605
368 GluLeuLeuTyrGlyLysThrSerProThrAlaGluThrIleLeuLys 384
2606 ACTATGCTCTCTCTCAAGGTCCCAAGACACAG..... 2641
384 rAsnIleSerSerGlyHisValProHisGlyProArgThrArgPro... 399
2642 .....CCAAAGACACTCCCTCTCA 2660
400 .....SerGluGlnLeuTyrTyrLeuSerArgAlaGlnGly 411
2661 CTGGCAGACACTTCATGACCAAGATGACCATGCTGTGAGC..... 2698
412 PheGlnValGluTyrLysAspPheProLysAsnAsnLysAsnGluCysVa 428
2699 .....CACCGGTGCTT 2709
428 lSerLeuIleAsnCysSerSerGlnProProLeuValSerHisGlyIleG 445
2710 CAACACTCTGACTAAGAGCTTCCAGCCCTCTGTCGCGGCAAGATTT. 2758
445 lylLysAspValGluSerCysHisAspMetAlaAlaLeuAsnIleLeuLys 461
2759 .....CTGGCCCGCATC...ATTATGAAA 2779

```

```

462 LeuLeuSerGluLeuAspGlnGlnSerThrGluMetProArgThrGlyAs 478
2780 AAAGACTCTGAC...GACATGGGTGTCGTGCTGACGCTTGGGAACAGGAA 2826
478 nglyProValSerAlaCysGlyArgCys 487
2827 T.....CGCTGT 2833

```

seq\_name: /cgn2\_6/plodata/2/1na/5A\_COMB.seq:US-08-457-459-1

seq\_documentation\_block:

; Sequence 1, Application US/08457459

; Patent No. 567428

; GENERAL INFORMATION:

; APPLICANT: Nishikura, Kazuko

; TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use

; NUMBER OF SEQUENCES: 39

; CORRESPONDENCE ADDRESS:

; ADDRESS: Howson and Howson

; STREET: Spring House Corporate Cntr, P.O. Box 457

; CITY: Spring House

; STATE: Pennsylvania

; COUNTRY: USA

; ZIP: 19477

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentln Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/457,459

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/197,794

; FILING DATE: 17-FEB-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/280,443

; FILING DATE: 25-JUL-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Bak, Mary E.

; REGISTRATION NUMBER: 31,215

; REFERENCE/DOCKET NUMBER: WST49CUSA

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 215-540-9206

; TELEFAX: 215-540-5818

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 6671 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: unknown

; MOLECULE TYPE: CDNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 155..3832

; US-08-457-459-1

alignment\_scores:

Quality: 171.00

Ratio: 0.710

Percent Similarity: 45.817

Percent Identity: 21.483

alignment\_block:

US-09-316-048-11 x US-08-457-459-1 ..

Align seg 1/1 to: US-08-457-459-1 from: 1 to: 6671

55 LysMetArgProProValLysHisAspAlaProAlaArgAla..... 68

..... 68

```

1445 AGACGTAAACACCTGTTACATTAACATGGCCCTCAAAAGCAGGGTATGT 1494
69 .....LeuA 70
1495 TGACTTTGAAATGGCCAGTGGGCCACAGATGATCCCGATGACTTGA 1544
70 rgtThrLeuGlnSerGluProLeuProGluArg.....LeuGluValAsn 84
1545 ATAGTATCCGCGCAGCAGCAGGTGATGTTTCGAGCCATCATGAGATGCCC 1594
85 Gly.....ArgGluValAsn 89
1595 TCCTTTTACATGATGCTGCTCCAGCGTGTTCACCTTACAAAGAACTGAC 1644
89 ugiGluAsnLeuAsnLysSerGluIleSerGluValPheGluIleAla 106
1645 AGAGTGGCCAGCTG...AGAAGCCCATTCACCGGGCTGTGATGATATGCC 1691
106 euLysArgAsnLeuProValAsnPheGluValAlaArgGluSerGlyPro 122
1692 AGTTCGCTGATCAACCTGTGAGTCAACATGATAGAGCAGAGTGGACCA 1741
123 ProHisMetLysAsnPheValThrArgValSerVal...GlyGluPheVal 138
1742 CCCCATGAACCTCGATTTAAATTCAGGTTTCATCATGCGCGAGAGTT 1791
138 lgiGluGluGlyGlyLysSerLysLysIleSerLysLysAsnAlaAla 154
1792 TCCCCAGCTGGAAGCTGGAAAGCAAGAAAGTGGCCAAAGCAGATGCACCTA 1841
155 .....ArgAlaValLeuGluGluLeuArgArgLeuProProLeu 167
1842 TGAAGCCATGACATTTCTGTAGAGAAAGCCAAAGCCAGAGACAGTGA 1891
168 ProAlaValGluArgValLysProArgLleLysLysSerGlnProth 184
1892 AAATGCAAGAAATCATCCCATTTCCACAGAGAAAGAAATCAGAGAAAGAC 1941
184 rCysLys.....ThrAlaProAspTyrGlyGlnGly 195
1942 TGCAGAGTCCCAAGACCCCAACCCCTTCAGCCACATCTTTTTCGCGGA 1991
195 eLysProIleSerArgLeuAlaGlnIleGlnGlnAlaLysLysGluLys 211
1992 AGAGCCCCCGTACACACACTGCTGAGTGTATGCAC...AAATGGGGAAC 2038
212 GluProGluTyrMetLeuLeuThrGluArgGlyLeuProArgArgArg 228
2039 TCCTGGCAATTCGCTCTCTGTCCAAGAGGCCCTGCCATGAACCCAA 2088
228 urPheValMetGlnValLysValGlyHisThrAlaGluGlyValGly 245
2089 GTTTCACATTAATCTGTTGCAAGTGGAGCCCAAACTTTCCCAAGTGTGAGT 2138
245 hr...AsnLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGlu 260
2139 CTCGCCAGCAAGAAAGTGGCAAAAGCAGATGGCCGACAGAGAACCATGAAG 2188
261 IleLeuGlyPheLysVal.....ProGlu 268
2189 GCCCTTCATGAGGAGGAGCCCAACTCATGCTTGTGATAACCACTGGA 2238
268 naLagInProAlaLysProAlaLeuLysSerGluGluLysThrProVal 265
2239 AGGTATGATGTCAGAGTCACTTATTAATCTGGAATCATGATGCCACACA 2288
285 yslLysProGluAspGlyArgLysValThrPhePheGluProSerPro... 300
2289 AGGTCAAGGAAGATTGGGAGACTGTGATGATACCTGAACACCAACCTGTG 2338
300 ..... 300
2339 GGTGGCTTTTGGAGTACGCCGCTCCATGCTTGTCTGTGAATTCAA 2388

```

```

301 ....GlyAspGluAsnGlyThrSerAsnLysAspGluGluPheArgMetP 316
3289 GTTGTGTGACACAGTCCGGACCTCCTCAGCAG.....C 2420
316 rGlyrLeuSerHisGlnGlnLeuProAlaGly...IleLeuProMetVal 331
2421 CCAAGTTCTGTTTACCAAGCAAAAGTTGGGGGTGCTGCTGCCAGCGCTC 2470
332 .....ProGluValAlaGlnAla..... 337
2471 TCGGCACACAGCAAGAAAGCAGACAGCAAGAGCAGATGGGCTCT 2520
338 .....ValGlyValSerGlnGlyHisThrLysAspPheThr 351
2521 CCGGTCTGATTTGGGAGAACGAGAGGAGCAAGCATGGGTTTCACAG 2570
351 rGAlaAlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArg 367
2571 AGGTAAACCCCA.....GTGACAGGGGCGCAGTCTCAGAGA 2605
368 GluLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLys 384
2606 ACTATGCTCTCTCTCAGAGTCCCGAAGACACAG..... 2641
384 rAsnIleSerSerGlyHisValProHisGlyProArgThrArgPro.... 399
2642 .....SerGluGlnLeuTyrTyrLeuSerArgAlaGlnGly 411
400 ..... 411
2661 CTGGCAGACCTTCATCCATGACCATGACCATGCTGAGC..... 2698
412 PheGlnValGluTyrLysAspPheProLysAsnAsnLysAsnGlyCysVa 428
2699 .....CACGGTCTT 2709
428 lSerLeuIleAsnCysSerSerGlnProLeuValSerHisGlyIleG 445
2710 CAACACTCTGACTACAGCTTCACGCCCTCTGCTGCGCCGCAAGATT. 2758
445 lLysAspValGluSerCysHisAspMetAlaAlaLeuAsnIleLeuLys 461
2759 .....CTGGCCGCGCATC...ATTATGAA 2779
462 LeuLeuSerGluLeuAspGlnGlnSerThrGluMetProArgThrGlyAs 478
2780 AAAGACTCTGAG...GACATGGGTGCTGCTGCTGAGCTTGGGAACAGGAA 2826
478 nglyProValSerAlaCysGlyArgCys 487
2827 T.....CGCTGT 2833

seq_name: /c9n2_6/plodata/2/lna/5A.COMB.seq:US-08-555-678-1
seq_documentation_block:
; Sequence 1, Application US/08555678
; Patent No. 5763174
; GENERAL INFORMATION:
; APPLICANT: Nishikura, Kazuko
; TITLE OF INVENTION: RNA Editing Enzyme and Methods
; TITLE OF INVENTION: of Use Thereof
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/555,678
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/197,794
FILING DATE: 17-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/280,443
FILING DATE: 25-JUL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/457,459
FILING DATE: 01-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: WST49DUSA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9206
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6671 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 155..3832
US-08-555-678-1

alignment_scores:
Quality: 171.00 Length: 526
Ratio: 0.710 Gaps: 23
Percent Similarity: 45.817 Percent Identity: 21.483

alignment_block:
US-09-316-048-11 x US-08-555-678-1 ..
Align seg 1/1 to: US-08-555-678-1 from: 1 to: 6671

55 LysMetArgProProValLysHisAspAlaProAlaArgAla..... 68
.....
1445 AACAGTGAACACCTGTTTATTACATGAGCCCTCAAAAGCAGGATATGT 1494
69 .....LeuA 70
1495 TGACTTTGAAATGGCAGTGGCCACACATGACATCCAGATGACTTGA 1544
.....
70 rgtThLeuGlnSerGluProLeuProGluArg.....LeuGlnValAsn 84
.....
1545 ATAGATATCGCGCAGCAGCAGGTGAGTTTCGACCATCATGAGATGCC 1594
85 Gly.....ArgGlnAlaGln 89
.....
1595 TCCTTCTACAGTCATGCTTGGCCAGGCTTTCACCTTACAGAAAGTGCAC 1644
89 uGluGlnAsnLeuAsnLysSerGlnIleSerGlnValPheGlnIleAlaL 106
1645 AAGATGCGCCAGCTG...AAGAACCCCATCAGCGGCGCTTAGAATATGCC 1691
106 euLysArgAsnLeuProValAsnPheGlnValAlaArgGlnSerGlyPro 122
1692 ACTTGCGTCAAGTCAAACTGTGATTCACATGATAGACAGAGTGGACCA 1741
123 ProHisMetLysAsnPheValThrArgValSerVal...GlyGluPheVal 138
1742 CCCCATGAACCTCGATTAAATTCACAGTTTGCATCATGATGCGGAGAGTT 1791

```

```

138 lGlyGlnGlyGlnGlyLysSerLysLysIleSerLysLysAsnAlaAla 154
.....
1792 TCCCCAGCTGAAGCTGGAAGCAAGAAAGTGGCCAAAGCAGATGCAAGTA 1841
155 .....ArgAlaValLeuGlnLeuArgArgLeuProLeu 167
.....
1842 TGAAGCCATGACAAATTCGTAGAGGAAGCCAAAGCAGACATGGA 1891
168 ProAlaValGluArgValLysProAlaArgIleLysLysSerGlnProth 184
1892 AATTCAGAAAGATTCATCCCATTTCCACAGAGAAAGATTCAGAGAAC 1941
184 rCysLys.....ThrAlaProAspTyrGlyGlnGly 195
1942 TGCAGAGTCCAGACACCCACCCCTTCCAGCAATCCTTTCTTGGGGA 1991
195 eLAsnProIleSerArgLeuAlaGlnIleGlnGlnAlaLysLysGlyLys 211
1992 AGAGCCCGCTCACACACACTGTGAGTGTATGCAC...AAATGGGGAAC 2038
212 GluProGluTyrMetLeuLeuThrGluArgGlyLeuProArgArgArg 228
2039 TCTCGCAATTCCTCGTCTCTCTCCAAAGAGGCCCTGCCCATGAACCCA 2088
228 uPheValMetGlnValLysValGlyHisIleThrAlaGluGlyValGly 245
2089 GTTCCAAATCTGTGTGAGTGGAGCCCAACTTTCCCACTGATGAGTG 2138
245 hr...AsnLysLysValAlaLysArgAsnAlaAlaGlnAsnMetLeuGlu 260
2139 CTCGCCAGCAAGAAAGTGGCAAAAGCAGATGCGCCAGAGGAAACCAATGAAG 2188
261 lLeuGlnLysPheLysVal.....ProGln 268
2189 GCCCTGCATGGGAGGAGCGGACCAATCCATGCTTCTGATACACAGCTTA 2238
268 nAlaGlnProAlaLysProAlaLeuLysSerGlnGluLysThrProValL 285
2239 AGTATGATCTCAGAGTCACTGATTAATCTTGAATCCATGATGACCAACA 2288
285 yLysProGluAspGluArgLysValThrPhePheGluProSerPro... 300
2289 AGTCAGAGAAATGGCGAGCTGTGATACCTGACCAACCAACCTGTG 2338
300 ..... 300
2339 GGTGGCCTTTGGAGTACGCCCGCTCCCATGCTTGTCTGATTAACATCA 2388
301 ...GlyAspGluAsnGlyThrSerAsnLysAspGluGluPheArgMetP 316
2389 GTTGTGTCAGACAGTCCGACCTCTCAGCAG...C 2420
316 rOtyrLeuSerHisGlnGlnLeuProAlaGly...lLeuProMetVal 331
2421 CCAAGTGTGTTTACCAAGCAAAAGTTGGGGTGTGCTGTTCACAGCGT 2470
332 .....ProGlnValAlaGlnAla... 337
2471 TGCAGCACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2520
338 .....ValGlyValSerGlnGlyHisIleThrLysAspPheThr 351
2521 CGGTGCTGTGATTTGGGAGAGACAGAGCAAGCAAGCAAGCAAGCAAG 2570
351 rGAlaAlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArg 367
2571 AGTAAACCCA.....GTGACAGGGCCAGTGTCAAGAA 2605
368 GluLeuLeuTyrGlyLysThrSerProThrAlaGluThrIleLeuLys 384
2606 ACTATGCTCTCTCTCAAGTCCCAAGCAAGCAAG... 2641
384 rAsnIleSerSerGlyHisValProHisGlyProArgThrArgPro... 399

```





2139 CTCGCCAGCAAGAAAGTGGCAAGACAGTATGGCCGAGAGCAAGCAATCAG 2188  
261 lIleuGllyPheIysVal.....ProG1 268  
||| :||| :|||  
2189 GCCCTCATGTGGGAGGAGGACCACTCCATGGCTTCTGATACCAAGCTGA 2238  
268 nIaGlnProAlaIlySerProAlaIleuIysSerGlnGluIlyThrProVal 285  
||||| :||| :|||  
2239 AGGTATGATCTCAGAGTCACTGTATATCTTGGAAATTCATGATGCCAACA 2288  
285 yAsySProG1IAspG1ArgIlyValIThrPheGlnProSerPro... 300  
||| :||| :||| :||| :||| :|||  
2289 AGGTCAAGAAAGATTTGGGACCTGTGTAGATACCTGAACACCACCCGTG 2338  
300 ..... 300  
2339 GGTGGCCTTTTGGAGTACGCCGCCCTCCATAGCTTGTCTGTGAATTCAA 2388  
301 ...G1AspG1IAsnG1IyThrSerAnIyAspG1GluIThrArgMetP 316  
||| :||| :||| :||| :||| :|||  
2389 GTTGTGTCGACCACTCCGACCTCTCTCAGAG.....C 2420  
316 rGtYIleuSerHISGlnIleuProAlaIly...IleuProMetVal 331  
||| :||| :||| :||| :||| :|||  
2421 CCAGATTCTGTTTCCAAAGAAAGTTGGGGGTCCCTGGTTTCCAGCCGTC 2470  
332 .....ProGluValAlaGlnAla..... 337  
||| :||| :||| :||| :||| :|||  
2471 TGGCCACACAGCAAGAAAGCAAGGCACAGACAGACAGTCCGGCTCT 2520  
338 .....ValG1yValSerGlnG1yHISIlyThrIyAspPheThr 351  
||| :||| :||| :||| :||| :|||  
2521 CCCTGTCCTGATTGGGAGAAACGAGAAAGCAAGCATGGGCTTTCACAG 2570  
351 rGAlaIAspProAlaIyAlaIyAlaIyThrAlaMetIleAlaArg 367  
||| :||| :||| :||| :||| :|||  
2571 AGGTAAACCCA.....GTGACAGGGGGCAGTCTAGAAAGA 2605  
368 G1uIleuIyTrG1yIyThrSerProThrAlaGluIThrIleuIysSe 384  
||| :||| :||| :||| :||| :|||  
2606 ACTATGCTCTCTCTCCAGAGTCCGCCAGAGCACAG..... 2641  
384 rAsnIleSerSerG1yHISValProHISG1yProArgThrArgPro... 399  
||| :||| :||| :||| :||| :|||  
2642 .....CCAAAGACACTCCCTCTCA 2660  
400 .....SerG1uIleuIyTrIyIleuSerArgAlaGlnG1y 411  
||| :||| :||| :||| :||| :|||  
2661 CTGGGACACACTTCCATGACCAAGATACCACTGCTGAC... 2658  
412 PheGlnValG1uIyTrIyAspPheProIyAsnAnIyIysAnG1ySva 428  
||| :||| :||| :||| :||| :|||  
2699 .....CACCGGTCTT 2709  
428 lSerIleuIleAsnCySerSerGlnProProIeValSerHISG1yIleG 445  
||| :||| :||| :||| :||| :|||  
2710 CAACACTCTGACTAAACAGCTTCCAGCCCTCTGCTGGCGGCAAGATT. 2758  
445 lYyAspValAlaIySerCyHISAspMetAlaIleuAnIleuIys 461  
||| :||| :||| :||| :||| :|||  
2759 .....CTGGCCGCCATC...ATTATGAA 2779  
462 lLeuIleSerG1uIleuAspGlnSerThrG1uMetProArgThrG1yAs 478  
||| :||| :||| :||| :||| :|||  
2780 AAAGACTCTAG...GACATGGGTGTCTGTCTAGCTTGGAAACAGGGA 2826  
478 nG1yProValSerAlaCyS1yArgCyS 487  
||| :||| :||| :||| :||| :|||  
2827 T.....CGCTGT 2833

```

: Sequence, Application US/08931299
: Patent No. 6043219
: GENERAL INFORMATION:
: APPLICANT: Iandolo, John J.
: APPLICANT: Crupper, Scott S.
: TITLE OF INVENTION: Broad Spectrum Chemotherapeutic Peptide
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Hovey, Williams, Timmons & Collins
: STREET: 2405 Grand Boulevard, Suite 400
: CITY: Kansas City
: STATE: Missouri
: COUNTRY: U.S.A.
: ZIP: 64108
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/931,999
: FILING DATE:
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/710,561
: FILING DATE: 19-SEP-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Collins, John M.
: REGISTRATION NUMBER: 26,262
: REFERENCE/DOCKET NUMBER: 25043-A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 816/474-9050
: TELEFAX: 816/474-9057
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6755 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Staphylococcus aureus
: STRAIN: UT0007
: US-08-931-999-4

alignment_scores:
      Quality: 129.00          Length: 379
      Ratio: 0.754             Gaps: 12
Percent Similarity: 45.119    Percent Identity: 20.580

alignment_block:
US-09-316-048-11 x US-08-931-999-4 ..
Align seg 1/1 to: US-08-931-999-4 from: 1 to: 6755

84 AsnGlyArgGluaIagLugLnLeuAnLySserGIUIlSerGI 100
   |||:::|||||:::|||||:::|||||:::|||||:::  ::::
126 AACCAACGCCACACCGCCGAGGAAGGAGAACCAAGCACACAAGAAMA 175
   |||:::|||||:::|||||:::|||||:::|||||:::  ::::
100 nValPheGluIleAlaLeuLyARgAsnLeuProValAsnPheGluVal. 116
   |||:::|||||:::|||||:::|||||:::|||||:::  ::::
176 AGAACCAGCAACAAAAACAGAAAGACAAACCAAAGAAAGAAAGCA 225
   |||:::|||||:::|||||:::|||||:::|||||:::  ::::
117 .....AlaArgGluSerGlyProProHisMetLysAsnPheValThrArg 131
   ::|||:::|||||:::|||||:::|||||:::|||||:::  ::::
226 ACCCAGAAAAGAGAGAGAGCAAGAGAAAGAGAGAGAGAGAGAGAGCA 275
   ::|||:::|||||:::|||||:::|||||:::|||||:::  ::::
132 ValSerValGlyGlnPheValGlyGlyGlyGlyGlyGlySerLysLys 148
   ::::|||:::|||||:::|||||:::|||||:::|||||:::  ::::
276 AAAGCAGAGGGCGAGACACAGGGAACAGCAAAAAAAAAAACAHAACCCAGCG 325

```



```

156 AlaValleuGlulnleuArgArgleuProleuProAlaValGluar 172
      :::::||||| :::::
394 TTGCGTGTGAGATCTTAAT ..... 414
172 gVallylProArgIlleuLysLysSerGlnProthrCysLysThrAlaP 189
      :::::||||| :::::
415 .....AAGAAAGAAAGAGCAGTTCCTTTATTATTGACAAACA 454
189 roAspTyrGlyGlnGlyMetAsnPro.....IleSer 199
      :::::||||| :::::
455 CGAATTCTTCAGAGATTAATCCATGGGCAATTACATAGCCCTTAATCAAT 504
200 ArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTyrMe 216
      :::::||||| :::::
505 AGAATTGGCCAG.....AAGAAAGACTAAGCTAAATATATGA 542
216 tLeuLeuThrGluArgGlyLeuProArgArgGluPheValMetGlnV 233
      :::::||||| :::::
543 ACAGTGTGCATCG...GGGCTGCATGGCCGCAAGAGATTCATTATTAAT 589
233 alLysValGlyHis...HisThrAlaGluGlyAlaGlyThrAsnLys 248
      :::::||||| :::::
590 GCAAAATGGGACAGAAAGAAATATATGTTGGTACAGGTCTCTAATAACAG 639
249 ValAlaLysArgAsnAlaAlaGlu...AsnMetLeuGluIleLeuGlyPh 264
      :::::||||| :::::
640 GAAGCAAAACAATTGGCCGCTAAACTTCATATCTCAGATATTA..... 664
264 eLysValProGlnAlaGlnProAlaLysProAlaLysSerGluGluL 281
      :::::||||| :::::
685 .....TCAGAAAG 691
281 ysthrProValLysLysPro....GlyAspGlyArgLysValThrPhe 295
      :::::||||| :::::
692 AAACCTCAGTAATCTGACTACCTGCTCTGTCTCTTCTTGTCTACTACG 741
296 PheGlu..... 297
      :::::||||| :::::
742 TGTGAGTCCCAAGCAACTTTAGTGACGACGACACATCGCTCTGAATC 791
298 .ProSerProGlyAsp.....GluAsnG 305
      :::::||||| :::::
792 ATCATCTGAAGGTGACTTCTCAGCAGATACATCAGATTAATTTAAACA 841
305 LysThrSerAsnLysAspGluGluPheArgMetProTyrIleuSerHisGln 321
      :::::||||| :::::
842 GTGACAGTTTAAACAGTCTTCTGTCCTTATGAAATGGTCTCAGAAATTAAT 891
322 Gln.....LeuProAlaGlyIleuLeuProMetVal 331
      :::::||||| :::::
892 CAAGAAGAGCAAAAGATCTTGGCACCACAGATTGACCTTCTCTGACAT 941
331 lProGluValAlaGlnAlaValGlyLysSerGlnGlyHisHisThrLysA 348
      :::::||||| :::::
942 GAAGAAGAAACAAGTATATCTGACAAAGAGTTTGGCATGATTTAAAG 991
348 sp.....PheThrArgAlaAlaProAsnPro 356
      :::::||||| :::::
992 AAATAGAAATTAATTGGCTCAGGTGATTTGGCCAGTTTCAAGCAAAA 1041
357 AlLysAlaThrValThrAlaMetIleAlaArgLysLeuLeuTyrGlyL 373
      :::::||||| :::::
1042 CACAGAAATTCAGCAAGAAAGACTTACGTTATTAAACGCTTAAATAT..... 1086
373 yThrSerProThrAlaGluThrIleLeuLysSerAsnIleSerSerGlyH 390
      :::::||||| :::::
1087 .AATAACGAGAAAGCGAGCGTGAAGTAAAGCATTTGGCAAACTGATGAT 1135
390 lValProHisGlyProArgThrArgProSerGluGlnLeuTyrTyrLeu 406
      :::::||||| :::::
1136 ATGTA.....AATATTGTTCACTACCAAT 1158

```

```

407 SerArgAlaGlnGlyPhe.....GlnVal 414
      :::::||||| :::::
1159 GCGTGTGGATGATGATTTGATATGATCCAGACAGCATGATGATTCCT 1208
414 lGluTyrLysAspPhe...ProLysAsnAsnLysAsnGluCysValSerL 430
      :::::||||| :::::
1209 TGAGACAGAGATTAATGATCTGAGAACAGCAAAATATGTTCAAGTCAA 1258
430 euLysAsnCys 433
      :::::||||| :::::
1259 AGACTAAGTCC 1269

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:us-08-434-998-8

seq_documentation_block:
: Sequence 8, Application US/08434998
: Patent No. 5866781
: GENERAL INFORMATION:
: APPLICANT: Silverman, Robert H.
: APPLICANT: Sengupta, Dibyendu N.
: TITLE OF INVENTION: Antiviral Transgenic Plants, Vectors,
: TITLE OF INVENTION: Cells and Methods
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &
: STREET: 200 E. Broward Boulevard
: CITY: Fort Lauderdale
: STATE: Florida
: COUNTRY: USA
: ZIP: 33301
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/434,998
: FILING DATE:
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/198,973
: FILING DATE: 18-FEB-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Manso, Peter J.
: REGISTRATION NUMBER: 32,264
: TELEPHONE/DOCKET NUMBER: C111363-16
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 305/527/2498
: TELEFAX: 305/764/4996
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2562 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-434-998-8

alignment_scores:
      Quality: 122.50      Length: 387
      Ratio: 0.677      Gaps: 17
      Percent Similarity: 46.770      Percent Identity: 22.481

alignment_block:
US-09-316-048-11 x US-08-434-998-8
Align seg 1/1 to: US-08-434-998-8 from: 1 to: 2562

107 LysArgAsnleuProValAsnPhelGluValAlaArgLysSerGlyPropr 123
      :::::||||| :::::
244 AACGAGGAGTACTACTTAATATCAAGAACTGCTTAATTCAGGACTCC 293

```

```

123  oHisMetLysAsnPhenValThrArgValSerVal...GlyLupPheValG 139
    ||||| ..... ||||| ..... ||||| .....
294  ACATGATGATGAGGTTTACATTTCAGATTATATGATGGAAGAAATTC 343
139  LylGluGlyGlyGlyLysSerLysLysLysLysLysLysLysLysLys 155
    ||||| ..... ||||| ..... ||||| .....
344  CAGAGAGTGAAGTATGATCAAGAAAGAAAGAAAGAAAGAAAGTCCG 393
156  AlValLeuGluGluGluGluGluGluGluGluGluGluGluGluGlu 172
    ..... ||||| .....
394  TTGAGCTGTGATGATCTTAAT..... 414
172  gValLysProArgLysLysLysSerGlnProThrCysLysThrAlaP 189
    :||| ..... ||||| .....
415  ..... AAGAAAGAAAGAGGAGTATGCTTTATTTATGCAACAA 454
189  roAspTyrGlyGlyGlyMetAsnPro.....LysSer 199
    :||| ..... ||||| .....
455  CGAATTCCTCAGAAAGATATCCATGGGAATTACATAGCCCTTATCA 504
200  ArgLeuAlaGlnIleGlnIleAlaLysGluLysGluProGluTyrMe 216
    ||||| ..... ||||| .....
505  AGAATTCGCCAG.....AAGAAAGACTAACTGTAATTTATGA 542
216  tLeuLeuThrGluArgGlyLeuProArgArgGluPheValMetGlnV 233
    :||| ..... ||||| .....
543  ACAGTGTGCATCG...GGGTGCATGGGCCAGAAAGATTCATATTAAT 589
233  allYsValGlyHis...HisThrAlaGluGlyValGlyThrAsnLys 248
    ||||| ..... ||||| .....
590  GCAAAATGGCAGAAAGAAATATAGTATGTTGACAGTTCCTACTAAAC 639
249  ValAlaLysArgAsnAlaAlaGlu...AsnMetLeuGluIleLeuGly 264
    ||||| ..... ||||| .....
640  GAAAGCAAAACAAATTGGCCGCTAACTTCATATTCAGATTTA..... 684
264  elYsValProGlnAlaGlnProAlaLysProAlaLeuLysSerGluGlu 281
    ..... ||||| .....
685  ..... TCAGAG 691
281  yThrProValLysLysPro.....GlyAspGlyArgLysValThrPhe 295
    :||| ..... ||||| .....
692  AAACCTCAGTAATCTGACTACCTGTCCTGCTCTTTGCTACTACG 741
296  PheGlu..... 297
    |||
742  TGTGATGCCAAAGCAACTCTTAGTGACAGACACTCGCTTCTGAATC 791
298  ProSerProGlyAsp.....GluAsnG 305
    ||| ..... ||||| .....
792  ATCATCTGAAGGTACTTTCACAGATACATCAGAGTAAATTTCTAACA 841
305  LylThrSerAsnLysAspGluGluPheArgMetProTyrLeuSerHisGln 321
    :||| ..... ||||| .....
842  GTGACAGTTTAAACAGTTCTCGTTGCTTGAATGTCGACAAATTAAT 891
322  Gln.....LeuProAlaGlyIleLeuProMetVal 331
    ||| ..... ||||| .....
892  CAAAGGAAGCAAAAGATCTTTGGCACCGAGTTTGACCTTCCTGACAT 941
331  lProGlnValAlaGlnAlaValLysSerGlnGlyHisHisThrLysA 348
    :||| ..... ||||| .....
942  GAAAGAAACAAGTATACTGTGACAGAGGTTTGGCATGGATTTTAAG 991
348  sp.....PheThrArgAlaAlaProAsnPro 356
    :||| ..... ||||| .....
992  AAATVAGATTAATTTGGCTCAGGTGATTTGGCAAGTTTCAAGCAAAA 1041
357  AlAlaLysAlaThrValThrAlaMetIleAlaArgLysLeuLeuTyrGly 373
    :||| ..... ||||| .....
1042  CACGAAATTGACGGAAGACTTACGTTATTAAGTGTAAATAT..... 1086

```

```

373  yThrSerProThrAlaGluThrIleLeuLysSerAsnLysSerSerGlyH 390
    :||| ..... ||||| .....
1087  AATAAGAGAGAGCGGAGCGCTGAAGTAAAGATTTGGCAAAACTTGATC 1135
390  lSValProHisGlyProArgThrArgProSerGluGlnLeuTyrTyrLeu 406
    ||||| ..... ||||| .....
1136  ATGTA.....AATATGTTCACCTAACAT 1158
407  SerArgAlaGlnGlyPhe.....GlnVal 414
    :||| ..... ||||| .....
1159  GCGCTTGGGATGATTTGATTATGATCCTGAGACAGTATGATTCCT 1208
414  lGluTyrLysAspPhe...ProLysAsnAsnLysAsnGluCysValSer 430
    :||| ..... ||||| .....
1209  TGACAGACAGTATGATGATCTCGAAGCAAGCAAAATAGTTCAAGGTCAA 1258
430  euLleAsnCys 433
    :||| ..... ||||| .....
1259  AGACTAAGTGC 1269

```

seq\_name: /cgn2\_6/prodata/2/ina/5B\_COMB.seq:US-08-487-797-8

```

seq_documentation_block:
; Sequence 8, Application US/08487797
; Patent No. 5866787
; GENERAL INFORMATION:
; APPLICANT: Silverman, Robert H.
; TITLE OF INVENTION: Transgenic Plants Co-Expressing A
; TITLE OF INVENTION: Functional Human 2-5a System
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Holland & Knight
; STREET: One E. Broward Boulevard, #1300
; CITY: Fort Lauderdale
; STATE: Florida
; COUNTRY: USA
; ZIP: 33301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,797
; FILING DATE: 07-JUNE-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Manso, Peter J.
; REGISTRATION NUMBER: 32,264
; REFERENCE/DOCKET NUMBER: C111363-16(C)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 305/468-7811
; TELEFAX: 305/463-2030
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2562 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-487-797-8

```

```

alignment_scores:
Quality: 122.50      Length: 387
Ratio: 0.677         Gaps: 17
Percent Similarity: 46.770  Percent Identity: 22.481

```

```

alignment_block:
US-09-316-048-11 x US-08-487-797-8 ..
Align seg 1/1 to: US-08-487-797-8 from: 1 to: 2562

```

```

alignment_scores:
seq_name: /sgn2_6/p/tdodata/2/ina/PCtUS_COMB.seq:PCR-US95-02058-8
seq_documentation_block:
; Sequence 8, Application PC/TUS9502058
; GENERAL INFORMATION:
APPLICANT: Silverman, Robert H.
APPLICANT: Sengupta, Dibyendu N.
TITLE OF INVENTION: Antiviral Transgenic Plants, Vectors,
NUMBER OF SEQUENCES: 11
CURRENT APPLICATION DATA:
ADDRESS: Ruden, Barnett, McClosky, Smith, Schuster &
ADDRESS: Russell
STREET: 200 E. Broward Boulevard
CITY: Fort Lauderdale
STATE: Florida
COUNTRY: USA
ZIP: 33301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/02058
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/198,973
FILING DATE: 18-FEB-1994
ATTORNEY/AGENT INFORMATION:
NAME: Manso, Peter J.
REGISTRATION NUMBER: 32,264
REFERENCE/DOCKET NUMBER: CL11363-16
TELECOMMUNICATION INFORMATION:
TELEPHONE: 305/527/2498
TELEFAX: 305/764/4996
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 2562 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
PCT-US95-02058-8

```

```
alignment_scores:
  Quality: 122.50      Length: 387
```

Ratio: 0.677 Gaps: 17  
Percent Similarity: 46.770 Percent Identity: 22.481  
alignment\_block:  
us-09-316-048-11 x PCT-US95-02058-8 ..

Align seg 1/1 to: PCT-US95-02058-8 from: 1 to: 2562

```

107 LysArgAsnLeuProValAsnPhelValAlaArgLysSerGlyPro 123
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
244 AACAGGAGATAGTACTTAATATCAAGAACTGCCTAATTCAGACCTCC 293
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
123 OHISmetLysAsnPhelValThrArgVal..GlyLupheValG 139
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
294 ACATAGATAGAGGTTTACATTTCAAGTTTATATAGATGAAGAATTTTC 343
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
139 LylGluGlyGluGlyLysSerLysLysLysSerLysLysAsnAlaAlaArg 155
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
344 CAGAAGGTGAGGTAGATCAAGAACAGAACCAAAAATGCGCGCCAAA 393
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
156 AlValLeuGluGluLeuArgArgLeuProLeuProAlaValGluAr 172
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
394 TTAGCTTTAGATACTTAAT..... 414
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
172 GValLysProArgLLeLysLysLysSerGlnProThrCysLysThrAlaP 189
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
415 .....AAGAAAGAGAGGAGGAGTTCCTTATTTATTCAGACAA 454
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
189 roAspTyrGlyGlnGlyMetAsnPro.....LisSer 199
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
455 CGAATTCCTGAGAAAGATATCCATGGGAATTACATAGGCGCTTATCAAT 504
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
200 ArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTyrMe 216
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
505 AGAATTCCTCCAG.....AAGAAAGAGTAACTGTAATTTATGA 542
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
216 LLeuLeuThrGluArgGlyLeuProArgArgArgGluPhelValMetGlnV 233
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
543 ACAAGTGCATCG...GGGTGCATGGCCAGAAAGATTTTCATATATAAT 589
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
233 allYsValGlyHis..HisThrAlaGluGlyValGlyThrAsnLysLys 248
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
590 GCAAAATGGGACAGAAAGAAATATATGATGTGTACAGGTCTCTACAAACAG 639
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
249 ValAlaLysArgAsnAlaAlaGlu..AsnMetLeuGluIleLeuGlyPh 264
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
640 GAAGCAAAACAAATTGGCCGCTAACTTCATATCTCAGATATTA..... 684
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
264 elYsValProGlnAlaGlnProAlaLysProAlaLeuLysSerGluGlu 281
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
685 .....TCAGAG 691
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
281 ySThrProValLysLysPro.....GlyAspGlyArgLysValThrPhe 295
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
692 AAACCTCAGTAATCTGACTACTGCTCTGTCTTTCTTTGCTACTACG 741
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
296 PhelGlu..... 297
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
742 TGTAGATCCCAAGCAACTTTAGTGACGACGACACTCGCTTGATC 791
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
298 ProSerProGlyAsp.....GluAsnG 305
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
792 ATCATCTGAAGGTACTTTCAGCAGATACATCAGAGTAAATTTCTACA 841
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
305 LylThrSerAsnLysAspGluGluPhelArgMetProTyrLeuSerHisGln 321
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
842 GTGACAGTTTAAACAGTTTCTGTTGCTATGATGATGCTCAGAAATAAT 891
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
322 Gln.....LeuProAlaGlyLLeuProMetVal 331
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
892 CAAAGGAGAGCAAAAGATCTTTGGACCAAGATTGACCTCTCGATCAT 941
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
331 lProGlnValAlaGlnAlaValGlyValSerGlnGlyHisHisThrLysA 348

```

```

942 GAAAGAAACAAAGATAGTACTGTGACAGAGGTTTGGCATGATTTTAAAG 991
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
348 sp.....PheThrArgAlaAlaProAsnPro 356
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
992 AAATAGATTAATTTGGCTCAGGTGGATTTGGCCAAAGTTTCAAGCAAAA 1041
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
357 AlAlaValaThrValThrAlaMetIleAlaArgLLeuLeuTyrGlyG 373
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1042 CACAGAAATTGACGGAAGAGCTTACGTATTAACGTGTTAAATAT..... 1086
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
373 YThrSerProThrAlaGluThrIleLeuLysSerAsnLisSerSerGlyH 390
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1087 AATAACGAGAGAGCGGAGCGTGAATAAAGCATTTGCCAAAACTTGATC 1135
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
390 lSValProHisGlyProArgThrArgProSerGluGlnLeuTyrTylLeu 406
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1136 ATGTA.....AATATTGTCACCTACAT 1158
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
407 SerArgAlaGlnGlyPhe.....GlnVa 414
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1159 GGCTGTGGATGGATTTGATATGATCCTGAGACAGTGTGATTCCT 1208
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
414 lGluTyrLysAspPhe..ProLysAsnAsnLysAsnGlnCysValSerL 430
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1209 TGAGACAGCATGATATGATCTCTGAGAACAGCAAAATAGTTCAAGTCAA 1258
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
430 euLisAsnCys 433
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1259 AGACTAAGTGC 1269
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:us-08-143-219-1
seq_documentation_block:
; Sequence 1, Application US/08143219
; Patent No. 5670330
; GENERAL INFORMATION:
; APPLICANT: Sosenberg, Nahum
; APPLICANT: Katze, Michael G.
; APPLICANT: Roy, Sophie
; APPLICANT: Koromilas, Antonis E.
; APPLICANT: Barber, Glen N.
; TITLE OF INVENTION: TUMOR-CELL ASSAY METHOD AND KIT
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 611 West Sixth Street
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM compatible
; OPERATING SYSTEM: PC-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/143,219
; FILING DATE: October 25, 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION DATA: described below
; APPLICATION NUMBER: 08/141,244
; FILING DATE: October 22, 1993
; APPLICATION NUMBER: 07/953,681
; FILING DATE: September 29, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Douglas E. Olson
; REGISTRATION NUMBER: 22,798
; REFERENCE/DOCKET NUMBER: 204/139
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600

```

two

```

: TELEFAX: (213) 955-0440
: TELEX: 67-3510
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2628 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: ORIGINAL SOURCE:
: INDIVIDUAL ISOLATE: HUMAN PKR GENE, FIGURE 5
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 187..1836
: US-08-143-219-1

```

```

alignment_scores:
  Quality: 121.50      length: 188
  Ratio: 1.191         Gaps: 8
  Percent Similarity: 54.255  Percent Identity: 27.128

```

alignment\_block:

US-09-316-048-11 x US-08-143-219-1 ..

Align seg 1/1 to: US-08-143-219-1 from: 1 to: 2628

```

107 LYSATGAsnleuProValAsnPhelValAlaArgGluSerGlyProP 123
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
244 AAGCAGGAGTACTACTTAATATCAAGAACTGCTAATTCAGACCTCC 293
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
123 OHSMetLysAsnPhelThrArgValSerVal..GlyGluPheValG 139
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
294 ACATGATAGAGGTTTACATTTCAAGTTAATAAGATGCAAGCAATTC 343
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
139 IyGluGlyGlyGlySerLysLysIleSerLysAsnAlaAlaArg 155
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
344 CAGAGGTGAGAGTATGATCAAGAGAGCAAGCAAAATGCGCAGCCAA 393
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
156 AlavalLeuGluGlnLeuArgGlyLeuProLeuProAlaValGluAr 172
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
394 TTAGCTGTGAGATCTTAT..... 414
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
172 GValLysProArgIleLysLysLysSerGlnProThrCysLysThra 189
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
415 .....AAGAAAGAGAGCAGCTTACTCTTATTATTGACACACAA 454
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
189 roAspTyrGlyGlnGlyMetAsnPro.....IleSer 199
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
455 CGAATTCCTCAGAAAGATTATCCATGCGGAATTAACATAGCCCTTA 504
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
200 ArgLeuAlaGlnIleGlnGlnAlaLysLysGluLysGluProGluTyr 216
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
505 AGAATTGCCCG.....AAGAAAGAGCTACTGTAATTAATATGA 542
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
216 tLeuLeuThrGluArgGlyLeuProArgArgArgGluPheValMetGln 233
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
543 ACAGTGTGATCG...GGGTGATGGCCACAGAGATTTCATTATAAAT 569
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
233 AlLysValGlyHis...HisThrAlaGluGlyValGlyThrAsnLys 248
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
590 GCAAAATGGAGCAAGAAATATAGTATTGTACAGTTCTACTAAACAG 639
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
249 ValAlaLysArgAsnAlaAlaGlu...AsnMetLeuGluIleLeuGly 264
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
640 GAAGCAAAACAATTGGCCGCTTAACCTGCATATCTTCAGATATTA.... 684
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
264 elysValProGlnAlaGlnProAlaLysProAlaLeuLysSerGluGlu 281
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
685 .....TCAGAG 691
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
281 ysthrProValLys 285
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

: ||| |||||
692 AAGCTCAGTGAAA 705

```

seq\_name: /cgn2\_6/plodata/2/ina/6\_COMB.seq:US-08-717-294-42

seq\_documentation\_block:

Sequence 42, Application US/08717294

Patent No. 6114148

GENERAL INFORMATION:

APPLICANT: SEED, BRIAN

APPLICANT: HAAS, JÜRGEN

TITLE OF INVENTION: HIGH LEVEL EXPRESSION OF

TITLE OF INVENTION: PROTEINS

NUMBER OF SEQUENCES: 110

CORRESPONDENCE ADDRESS:

ADDRESSEE: Clark & Elbing LLP

STREET: 176 Federal Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02110

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/717,294

FILING DATE: 20-SEP-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Elbing, Karen L

REGISTRATION NUMBER: 35,238

REFERENCE/DOCKET NUMBER: 00786/345001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-428-0200

TELEFAX: 617-428-7045

TELEX:

INFORMATION FOR SEQ ID NO: 42:

SEQUENCE CHARACTERISTICS:

LENGTH: 4451 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-08-717-294-42

```

alignment_scores:
  Quality: 117.00      length: 522
  Ratio: 0.539         Gaps: 30
  Percent Similarity: 41.571  Percent Identity: 22.031

```

alignment\_block:

US-09-316-048-11 x US-08-717-294-42 ..

Align seg 1/1 to: US-08-717-294-42 from: 1 to: 4451

```

31 ProPheProValPro.....ProLeuLeuTyrGlnValG 42
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
167 CCGCTTCCCCCGCGGTGCGCCCAAGAGCTTCCCTTCAACACACAG.... 211
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
42 uLeuSerValGlyGlnGlnPheAsnGlyLysGlyLysMetArgProp 59
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
212 .....CGTGGTGTACAGAAACCCGTGTGCGAGTTCAACGACAC 254
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
59 roValLysHis.....AspAlaProAla..... 66
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
255 CTGTTCACATTCGCAAGCGCGCGCCCGCTGTGATGGCTGTGGGCC 304
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```





## alignment\_block:

US-09-316-048-11 x US-09-335-409-1/rev ..

Align seg 1/1 to reverse of: US-09-335-409-1 from: 1 to: 68750

```

8 HisSerArgMetGlnSerThrTySerTyrGlyMetArgGlyValAlaTyr 24
   ||| |||||
   ||| |||||
51360 CACTTGAGGATGCGCATGCTCTCGTGGCA...AGCGTACCGCGCGAGCACC 51314
   ||| |||||
24 rProProAlaGlyTyrPheTyrProPheProValProProLeuLeuTyrGlnVal 41
   ||||| |||||
51313 TCCTCCGTCGCCGAGCCACACACTGCTCCACGCCCT..... 51280
   ||||| |||||
41 aGluLeuSerValGlyGlyGlnGlnPheAsnGlyLysGlyLysMetArg 57
   ||||| |||||
51279 .....GGAGCAGGAGCATATGATGAGGGCTGCTGCT 51247
   ||||| |||||
57 ..... 57
51246 TGAGCAGCGATAGAGCGCTCCGACGACGAGCCTCGACATCGCGT 51197
58 .....ProProValLysHisAspA 64
   ||||| |||||
51196 TGAATCTCCACGAGACTTCCTTCAACGCTCCACCGCTGAGCTGAGCC 51147
64 lAProAlaArgAla...LeuArgThrLeuGlnSerGluProLeuProGlu 79
   ||||| |||||
51146 GCTACACAGCTGCCCGTGGCGTGGCGCAGCCAGCTCTTACCTGCTCCT 51097
80 ArgLeuGluValAsnGlyArgGlnAlaGlnGluGluAsnLeuAsnLysS 96
   ||||| |||||
51096 CACGACTCGATCTGAAATGAGGCCGAGCCGACCTTCTTGGCTGAGG 51047
96 rGluLysSerGlnValPheGluLeuAlaLeuLysArgAsn..... 109
   ||||| |||||
51046 GCCACTTGACAGATGGCTCGCGCTCGAGGGCAGCGCGGCTCGCTC 50997
110 .....LeuProValAsnPheGluValAla 117
   ||| |||
50996 GAGCGGCAGCTGCTCCAGCACCGCTCGCGCTGCATAGAGATCTAGC 50947
118 ArgGluSerGlyProPro...HisMetLysAsnPheValThrArgValS 133
   ||||| |||||
50946 CGCGCGCAGAGCGCATCTCTACATACGCGCGCTGGCAACAGCGCTCG 50897
133 rValGlyGluPheValGlyGlnGlyLysSerLysLysIleSerL 150
   ||||| |||||
50896 CCATGAGCCTATGTTCCGAAAGGT.....AGCATAGCCT 50862
150 yslYsAsnAlaAlaArgAlaValLeuGlnGlnLeuArgArgLeuProP 166
   ||||| |||||
50861 CTGCTGCTCAGCGCTTGGCTCCACAAAGTGAAGCTC...CGGCTCGCGCG 50815
167 LeuProAlaValAlaGluArgValLysProArgLysLysLysSerGlnP 183
   ||||| |||||
50814 AGACGAGCAATGAG.....CCCCGAGCAGCGAGTGTCCGTGGCG 50774
183 oThrCysLysThrAlaProAspTyrGlyGlnGlyMetAsn..... 196
   ||||| |||||
50773 CCCTTCCGAAGTCCGAAAGCCGACGAGGCTTCGACCATGACAGCT 50724
197 ..ProLysSerArgLeuAlaGlnIleGlnGlnAlaLysLysGlySglu 212
   ||||| |||||
50723 GTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 50674
213 ProGluTyrMetLeuLeuThrGluArgGlyLeuProArgArgArgGluPh 229
   ||||| |||||
50673 CCC.....GCTGCGCGCGCGCGGGA..... 50653
229 eValMetGlnValLysValGlyHisHisThrAlaGlnGlyValGlyThr 246
   ||||| |||||
50652 .....ACAGCGAGCCAGCTCAGCGG 50631

```

```

246 snLysLysValAlaLysArgAsnAlaAlaGluAsnMetLeuGluLeu 262
   ||||| |||||
50630 ATAGCCGAGCGCCACAGCGTCCCGCAGCGCT..... 50599
263 GlyPheLysValProGlnAlaGlnProAlaLysProAlaLeu.....Ly 277
   ||||| |||||
50598 .....CCAGCAGCGTCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 50561
277 sSerGlnGluLysThrProValLysProGlyAspGlyArgLysValT 294
   ||||| |||||
50560 GAGCCACCGCAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 50512
294 hPhePheGluProSerProGlyAspGlyLysValLysValLysVal 310
   ||||| |||||
50511 .....GGGGGCGACGAGATGCGGTGCGGGCTCATCT..... 50479
311 GluGluPheArgMetProTyrLeuSerHisGlnGlnLeuProAlaGly 327
   ||||| |||||
50478 .....CGATGACAGCGCGCGCGCGCGCT.....CCAGCAGCGCT 50444
327 eLeuProMetValProGluValAlaGlnAlaValGlyVal.....Serg 342
   ||||| |||||
50443 TCGCGCGCGCGCAGCGAGCGCAGCGCTGCCAGAGTTGTCGCCCGCAGTA 50394
342 lngLysHisThrLysAspPheThrArgAlaAlaProAsnProAlaLys 358
   ||||| |||||
50393 GCTGCGCAGCGAGCTCGGACCGCGCATCACCGCGCGCTGCTGCTGCTGCT 50344
359 AlaThrValThrAlaMetIleAlaArgGluLeuTyrGlyGlyThrS 375
   ||||| |||||
50343 GCATGCG.....GCACCGCAGCGCTGCGCGCGGATGCGCT 50309
375 rProThrAlaGluThrIleLeuLysSerAsnLysSerGlyHisValP 392
   ||||| |||||
50308 CCAGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 50265
392 roHisGlyProArgThrArgProSer 400
   ||||| |||||
50264 GCTATGCTGCGGAGCTCCACCTTCA 50239

```

---





```

|||||
153 CGCAATGACGCGGAGAAACATGCTGGAGATCTTGTTCAAAATCCCGCA 202
268 nAAGlnProAlaLysProAlaLeuLysSerGluLysThrProVal 285
203 GGGCGACGCCACCAAAACCCGACATCAAGTCAGAGAGAAACACCATTA 252
285 yAspGlyAspGlyArgLysValThrPhePheGluProSerProGly 301
253 AGAAACCGAGGATGAGAGAAAGTAACCTTTTGAACCTGCTCTGG 302
302 AspGlnAspGlyThrSerAsnLysAspGluLysPheArgMetProTyr 318
303 GATGAATAAGGCGACTGATTAATAAGAGATGACTGAGATGCTTATCT 352
318 uSerHisGlnGlnLeuProAlaGlyIleLeuProMetValProGluVal 335
353 AAGTCATCAGACAGCTGCTGCTGGAATCTTCCCATGTCGCCGAGGTG 402
335 lAglnAlaValGlyValSerGlnGlyHisHisThrLysAspPheThrArg 351
403 CCCAGGCTGAGGAGTTAGTCAAGACATCACCAAGATTTTACCAG 452
352 AlaAlaProAsnProAlaLysAlaThrValThrAlaMetIleAlaArg 368
453 GCGAGCTCCGATCTCTCCAAAGCGCAGGTACTGATGATGATGATGAT 502
368 uLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysSer 385
503 GTTGTGTTATGGGGGACCTCGCCACAGCGCAGACCATTTTAAAGAT 552
385 snIleSerSerGlyHisValProHisGlyProArgThrArgProSerGlu 401
553 ACATCTCTTACGCGCACGTACCCATGACCTCTCACGAGACCTCTGAG 602
402 GlnLeuTyrTyrLeuSerAlaGlnGlyPheGlnValGluTyrLysAs 418
603 CAAGCTGAGCTATCTTCCAGAGTCAGAGGATTCAGAGTTGAATACAA 652
418 pPheProLysAsnAsnLysAsnGluCys...ValSerLeuIleAsn 434
653 CTTCCTCCCAAAATCACCAAGAACGAATTCGTATCTTATCAATGGCT 702
434 eSerGlnProProLeuValSerHisGlyIleGlyLysAspValGlu 450
703 CCTCTGAGCAGCTCTGATCCACATGATTCGCAAGGTGTCGGAATC 751
450 rCysHis...AspMetAlaLeuAsnIleLeuLysLeuLeuSerGlu 465
752 CTGGCAATGATGATGATGGTGGCTGGACATCTTAAGATGGCTGGGGA 801
466 LeuAspGlnGlnSerThrGluMetProArgThrGlyAsnGlyProVal 482
802 GTGGACCCCAAGGTCAAAATAATCCAGAAACGGAGGAGACCATGTT 851
482 er 482
852 CG 853
seq_name: gb_est43:AW320695
seq_documentation_block: 633 bp mRNA EST 25-JAN-2000
DEFINITION uc22b10.y1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:2609851 5'
similar to TR:095793 095793 STRAUFEN PROTEIN. // mRNA sequence.
ACCESSION AW320695
VERSION AW320695.1 GI:6750239
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 633)

```

```

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:
www-bio.lnl.gov/bbrp/image/image.html
MGI:1023711
Seq primer: -40RP from Glibco
High quality sequence stop: 356.
FEATURES
Location/Qualifiers
1..633
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image="IMAGE:2609851"
/clone_lib="NCI_CGAP_Mam6"
/sex="Female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1. SalI;
Site:2. NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT 176 a 167 c 163 g 127 t
ORIGIN
alignment_scores:
Quality: 1006.00 Length: 205
Ratio: 5.005 Gaps: 0
Percent Similarity: 97.573 Percent Identity: 93.204
alignment_block:
US-09-316-048-11 x AW320695 ..
Align seg 1/1 to: AW320695 from: 1 to: 633
268 GlnAlaGlnProAlaLysProAlaLeuLysSerGluLysThrProVal 284
|||||
16 CAGGCGCAGGCTGCGCAAGCAGCAGCATCAATCAGAAAGAGACTCCAGT 65
284 lLysLysProGlyAspGlyArgLysValThrPhePheGluProSerProG 301
|||||
66 AAAGAAACCGAGGAGAGGAAAGTAAGCTTTTGAACCTAGGCCCTG 115
301 lYAspGluAsnGlyThrSerAsnLysAspGluLysPheArgMetProTyr 317
|||||
116 GGGATGAATAATGGAAGTACTAGTAAAGAGACGAGAGATTCAGATGCTTAT 165
318 LeuSerHisGlnGlnLeuProAlaGlyIleLeuProMetValProGluVal 334
|||||
166 CTGAGCGATCAGCAGCGTCCAGCTGGAATTCCTCCCATGATGTCGGAAGT 215
334 lAlaGlnAlaValGlyValSerGlnGlyHisHisThrLysAspPheThr 351
|||||
216 TGGCCAGGCTGTCGGGTAGTCAAGGACCCACACCAAGATTTACCA 265
351 rGlnAlaLysProAsnProAlaLysAlaThrValThrAlaMetIleAlaArg 367
|||||
266 GGGCAGCTCCAAATCTCTCCCAAGGCAACGGTAACTGATGATAGCCCA 315
368 GluLeuLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysSe 384
|||||

```

316 GACGTGTTGACGGGGGACCTCGCCACAGCCGAGACCAATTTTAAAG 365  
384 rAsnIleSerSergIyHisValProHisGlyProArgThrArgProSerg 401  
|||||  
366 TAACATCTCTTCAGGCCAGCTACCCGATGGACCTGCACCTAGACCCCTG 415  
401 lucIuLeuTyrrLeuSerArgAlaInglIyPheGlnValGluTyrrLys 417  
|||||  
416 AGCAAGCTACTACTCTTCAGAGCCAGGAGTCCAGGTGAAATACAAA 465  
418 AspPheProLysAsnAsnLysAsnGluCysValSerLeuIleAsnGlySe 434  
|||||  
466 GATGTCTCCACAGACACAGAGAGAGTGTGTCTCATCAACTGCTC 515  
434 rSergGlnProProLeuValSerHisGlyIleGlyLysAspValGluSerg 451  
|||||  
516 CTCACACCCGCTCTCTCATGTCATGTCATCGCAAGAGATGTGAAGTCT 565  
451 ySHisAspMetAlaAlaLeuAsnIleLeuLysIleLeuSergIuLeuAsp 467  
|||||  
566 GTCATGATATGAGCTGCTGACATGTAAAGCTGTGTGAGTGTGAC 615  
468 GlnGlnSerThrGluMet 473  
:::|||||  
616 TCACAGAACACAGATATT 633  
seq\_name: gb\_est7:BE894295

seq\_documentation\_block: 752 bp mRNA 29-SEP-2000  
LOCUS BE894295 601437666P1 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:3922529 5',  
DEFINITION mRNA sequence.

ACCESSION BE894295  
VERSION BE894295.1 GI:10356518  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 752)  
AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: ATCC/DC/DTP  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLCM744 row: d column: 18  
High quality sequence stop: 654.

FEATURES  
source  
1..752  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3922529"  
/clone\_lib="NIH\_MGC\_72"  
/tissue\_type="melanotic melanoma"  
/lab\_host="PH10B (phage-resistant)"  
/note="Organ: SKIN; Vector: PCMV-SPORE; Site:1: NCI;  
Site\_2: Sall; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2 kb. Library constructed by Life  
Technologies."

BASE COUNT 223 a 194 c 192 g 143 t  
ORIGIN

alignment\_scores: 981.50 Length: 237  
Quality:

Ratio: 4.461 Gaps: 3  
Percent Similarity: 92.827 Percent Identity: 84.810  
alignment\_block:  
US-09-316-048-11 x BE894295 ..

Align seq 1/1 to: BE894295 from: 1. to: 752

191 TyrIleGlnGlyMetAsnProIleSerArgLeuAlaInglIleGlnAla 207  
|||||  
3 TATGGCCAGGGGATCATTCGATTAGCCGACTGCGCCAGATCCAGACG 52  
207 aLysGlyGluLysGluProGluTyrrMetLeuLeuThrGluArgLysLeu 224  
|||||  
53 AAAAAGAGAGAAGAGCCAGAGTACAGCTCTCCACAGAGGAGGCTCC 102  
224 roAlgaTgaTgGluGluPheValMetGlnValLysValGlyHisHisThrAla 240  
|||||  
103 CGCCCGCAGGAGTGTGTGATGCAAGTGAAGGTTGGAAACACACTGCA 152  
241 GluGlyValGlyThrAsnLysLysValAlaLysArgAsnAlaAlaGluAs 257  
|||||  
153 GAAGGAACGGGCACCAACAGAGAGTGGCCAAAGCCAAATGCACCGAGAA 202  
257 nMetLeuGluIleLeuGluLysPheLysValProGlnAlaGlnProAlaLysP 274  
|||||  
203 CATGCTGGAGATCCTTGTTCAAAAGTCCCGCAGGCGGACGCCAAC 252  
274 roAlaLeuLysSergLysLysThrProValLysLysProGlyAspGly 290  
|||||  
253 CCGCACTCAAGTCAAGAGAGAGAACACCCATAAAGAAACACAGG. GATGGA 301  
291 ArgLysValThrPhePheGluProSerProGlyAspGluAsnGlyThrSe 307  
|||||  
302 AGAAAGATACCTTTTGAACCTGCTGCGGAGTGAATAAGGACTAG 351  
307 rAsnLysAspGluGluPheArgMetProTyrrLeuSerHisGlnGluLeuP 324  
|||||  
352 TAATAAGAGGATGATGATCAGAGTCCCTTATCTAAGTATCAGCAGCTGC 401  
324 roAlaGlyIleLeuProMetValProGluValAlaGlnAlaValGlyVal 340  
|||||  
402 CTGCTGGAATCTTCCATGCTGCGCCGAGGTGCGCCAGGCTGTAAGAGATT 451  
341 SerGlnGlyHisHisThrLysAspPheThrArgAlaAlaProAsnProAl 357  
|||||  
452 AGTCAGAGACATCACACCAGAAAGATTTTACAGAGACGCTCGAATCTCTG 501  
357 aLysAlaThrValThrAlaMetIleAlaArgLyuLeuLeuTyrrGlyLys 374  
|||||  
502 CAAGGCCACGGTAACTGCCATGATAGCCCGAGAGTTGTAGTATGGGGCA 551  
374 hrSerProThrAlaGluThrIleLeuLys.SerAsnIleSerSergLysI 390  
|||||  
552 CCTGGCCACAGCCGAGATCATTTTAAAGATAAATCAATCTCTTCAGGCCA 601  
390 sValProHisGlyProArgThrArgPro...SergLyuIleuTyrrL 406  
|||||  
602 CGTACCCCATGACTCTCACAGATCTCTGTAGCAATGATATCTTCTTC. 650  
406 euserArgAlaGlnGlyPheGlnValGluTyrrLysAspPheProLysAsn 422  
|||||  
651 .....CAGAGTCCAGAGATTCGGGGTGTATACAAAGACTTCCCAAAAAA 695  
423 AsnLysAsn 425  
|||||  
696 CAGAAAGAT 704  
seq\_name: gb\_est74:BE674157

seq\_documentation\_block: 616 bp mRNA 08-SEP-2000  
LOCUS BE674157 7676a04.x1 NCI\_CGAP\_lu24 Homo sapiens cDNA clone IMAGE:3278862 3',  
DEFINITION

similar to TR:095793 095793 STAUFEIN PROTEIN. ; mRNA sequence.  
 BE674157  
 VERSION BE674157.1 GI:10034698  
 KEYWORDS  
 SOURCE

ORGANISM human.  
 Homo sapiens

REFERENCE  
 AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Eukaryota; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)

JOURNAL  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL, send email to:  
 info@image.llnl.gov  
 Seq primer: -40UP from Gldco  
 High quality sequence stop: 469.

FEATURES  
 source  
 Location/Qualifiers  
 1..616  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3278862"  
 /clone\_lib="NCI-CGAP.Lu24"  
 /tissue\_type="carcinoid"  
 /lab\_host="DH10B"  
 /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a  
 modified polylinker; Plasmid DNA from the normalized  
 library NCI-CGAP\_Lu5 was prepared, and ss circles were  
 made in vitro. Following HAP purification, this DNA was  
 used as tracer in a subtractive hybridization reaction.  
 The driver was PCR-amplified cDNAs from a pool of 5,000  
 clones made from the same library (clonoids  
 141920-141791 and 1520904-1522439). Subtraction by Bento  
 Soares and M. Fatima Bonaldo."  
 BASE COUNT 179 a 153 c 139 g 141 t 4 others  
 ORIGIN

alignment\_scores:

Quality: 959.00 Length: 202  
 Ratio: 4.969 Gaps: 0  
 Percent Similarity: 95.545 Percent Identity: 90.099

alignment\_block:

US-09-316-048-11 x BE674157 ..  
 Align seg 1/1 to: BE674157 from: 1 to: 616

```

281 LysThrProValLysLysProGlyArgLysValThrPhePheG 297
|||||
8 AAGACACCCCTTAAGCAACCGAGGATGGAAGAAAGTAACCTTTTGA 57
|||||
297 uProSerProGlyAspGluAsnGlyThrSerAsnLysAspGluGluPheA 314
|||||
58 ACCTGGCTCGGGGATGAAATGGACTACTAATAAGAGATGATTC 107
|||||
314 rGneProTyrLeuSerHisGlnGlnLeuProAlaGlyIleuPromet 330
|||||
108 GATGCTTATCTAAGTCATCAGACGCTGCTCGGAATTTCTCCATG 157
|||||
331 ValProGluValAlaGlnAlaValGlyValSerGlnGlyHisThrI 347
|||||
158 GTGCCCGAGGTGCCCGAGGCTGTAGAGATTAGTCAAGACATCACAC 207

```

```

347 sAspPheThrArgAlaAlaProAsnProAlaLysAlaThrValThrAlaM 364
|||||
208 AGATTTTACAGGCGAGCTCCGAAATCCTCGCAAGGCGACGATTAAGTCCA 257
|||||
364 etIleAlaArgGluLeuLeuTyrGlyGlyThrSerProThrAlaGluThr 380
|||||
258 TGATACCCCGAGAGTGTGTGATGAGGGGCGACCTCGCCACAGCGAGACC 307
|||||
381 IleLeuLysSerAsnIleSerSerGlyHisValProHisGlyProArgTh 397
|||||
308 ATTTTAAGATATACATCTCTTCAGGCCACGTACCCCATGAGACTCTCAC 357
|||||
397 rArgProSerGluGlnLeuTyrTyrLeuSerArgAlaGlnGlyPheGlnV 414
|||||
358 GAGACCCCTGTGAGCACTGTGACTATCTTTCAGAGTCAGAGATTCAGG 407
|||||
414 aGluTyrLysAspPheProLysAsnAsnLysAsnGlyCysValSerLeu 430
|||||
408 TTGAATATCAAGAGACTTCCCAAAACACAAAGAACGATTTGATCTCTT 457
|||||
431 IleAsnCysSerSerGlnProProLeuValSerHisGlyIleGlyLysAs 447
|||||
458 ATCAATGCTCTCTTCANCCACCTGTGATCAGCATGATGATCAGCAAGA 507
|||||
447 pValGluSerCysHisAspMetAlaAlaLeuAsnIleLeuLysLeuLys 464
|||||
508 TGTGAGATCCTGCGCATGATATGCTGCGGTGACATCTTAAGATTCGTG 557
|||||
464 erGluLeuAspGlnGlnSerThrGluMetProArgThrGlyAsnGlyPro 480
|||||
558 CTGAATAGTGACCAACAANTACAGAGATGCCAAGAACANAGACACCA 607
|||||
481 ValSer 482
608 ATGCTCT 613

```

seq\_name: gb\_est73:BE613972

seq\_documentation\_block: 686 bp mRNA

LOCUS BE613972 EST 24-AUG-2000  
 DEFINITION 601504093F1 NIH\_MGC\_71 Homo sapiens cDNA clone IMAGE:3905783 5',  
 mRNA sequence.

ACCESSION BE613972  
 VERSION BE613972.1 GI:9895569  
 KEYWORDS EST.

SOURCE  
 ORGANISM human.  
 Homo sapiens

REFERENCE  
 1 (bases 1 to 686)  
 Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Eukaryota; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LCM700 row: j column: 24  
 High quality sequence stop: 675.

FEATURES  
 source

Location/Qualifiers  
 1..686  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3905783"  
 /clone\_lib="NIH\_MGC\_71"  
 /tissue\_type="leukomyosarcoma"  
 /lab\_host="DH10B (phage-resistant)"



/note="Organ: uterus; Vector: PCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo df.  
Average insert size 2.1 kb."

BASE COUNT 226 a 156 c 196 g 108 t  
ORIGIN

alignment\_scores:  
Quality: 915.00 Length: 230  
Ratio: 4.378 Gaps: 4  
Percent Similarity: 90.870 Percent Identity: 84.348

alignment\_block:  
US-09-316-048-11 x BE613972 ..

Align seg 1/1 to: BE613972 from: 1 to: 686

```

62 HisapalaprAlaArgAlaLeuArgThrLeuGlnSerGluProLeuP 78
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3 CACGATGCTGCTGCCAAGCGTTGAGATCTGCAGATGAGCCCTGCC 52
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
78 oclArgLeuValAsnGlyArgGluAlaGluGluGluAsnLeu 95
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
53 AGAGAGCTGGAGTGAAGAGATCCGAGAGAGAGAGAGATCAATCA 102
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
95 ysserGluLeSerGlnValPheGluLeuAlaLeuLysArgAsnLeuP 111
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
103 AATCTGAATTAAGTCAAGTGTGAGATGACATTAACGCACTTGCT 152
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
112 ValasnPheGluValAlaArgGluSerGlyProPheHisMetLysAsnPh 128
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
153 GTGAATTCGAGGTGGCCGGAGAGTGGCCACCCACATGATGAAGACTT 202
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
128 eValThrArgValSerValGlyGluPheValGlyGluGlyLys 145
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
203 GTGACCAAGGTTTCGTGGGGA...GTTGTGGGGAGGTGAAGGAGAA 249
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
145 eValLysValSerLysAsnAlaAlaArgAlaValLeuGluGlnLeu 161
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
250 GCAGAGATTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 299
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
162 ArgArgLeuProPheProLeuProAlaValGluArgValLysProArgLe 178
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
300 AAGAAGTATCGCCCTCTGCTGAGTTGAAGAGTAAAGCTAGAAATCA 349
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
178 sLysLysSerLeuProThrCysLys.....ThrAlaProAspTyrGly 193
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
350 AAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 399
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
193 InGlyMetAsnProLysSerArgLeuAlaGlnLeuGlnAlaLysLys 209
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
400 AGGATTC.AATCCGATTAGCCGACTGGCCCAAGATCCAGAGCAAAAAG 448
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
210 GluLysGluProGluTyrMetLeuLeuThrGluArgGlyLeuProArg 226
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
449 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 498
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
226 GArgGluPheValMetGlnValLysValGlyHisHisThrAlaGly 243
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
499 CAGGAGAGTTGTGATGACAGGTGAAGTTGAAACCACTGCAGAGGAA 548
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
243 aLgLYThrAsnLysLysValAlaLysArgAsnAlaGlnAsnMetLeu 259
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
549 CGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 598
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
260 GluLeuLeuGly.PheLysValProGlnAlaGlnProAlaLysProAla 276
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
599 GAGATCTCTGTTCAAAAGTCCCGAGAGAGAGAGAGAGAGAGAG 645
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
276 eLysSerGluGluLysThrProValLysLysProGly 288
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
646 TCAAGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 683

```

seq\_name: gb\_est40:AW141293

seq\_documentation\_block:

LOCUS AW141293 504 bp mRNA EST 30-OCT-1999  
DEFINITION EST291334 Normalized rat brain, Bento Soares Rattus sp. cDNA clone  
RGIBK57 5' end similar to maternal effect protein STRAUFEN, mRNA

sequence.

ACCESSION AW141293  
VERSION AW141293.1 GI:6161078

KEYWORDS EST.

SOURCE Rattus sp.

ORGANISM Rattus sp.

REFERENCE 1 (bases 1 to 504)

AUTHORS Lee,N.H., Glodex,A., Chandra,I., Mason,T.M., Quackenbush,J.,

TITLE Kerlavage,A.R. and Adams,M.D.

JOURNAL Rat Genome Project: Generation of a Rat EST (RESt) Catalog & Rat

COMMENT Unpublished (1998)

ATCC Contact: Lee, NH

The Institute for Genomic Research

9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3529

Fax: (301)-838-0208

Email: nhlee@tigr.org

For clone availability, additional sequence and expression

information related to this EST please check the TIGR Rat Gene

index (<http://www.tigr.org/tdb/rgi.html>). To order a clone

contact the ATCC (<http://www.atcc.org/atcc.html>).

Seq primer: M13 Reverse.

FEATURES

Location/Qualifiers

1..504

/organism="Rattus sp."

/db\_xref="taxon:10118"

/clone="RGIBK57"

/note="Organ: brain; Vector: pT733pac; Site\_1: EcoRI;

Site\_2: NotI"

BASE COUNT 138 a 143 c 117 g 106 t

ORIGIN

alignment\_scores:

Quality: 854.00 Length: 167

Ratio: 5.145 Gaps: 0

Percent Similarity: 99.401 Percent Identity: 97.006

alignment\_block:  
US-09-316-048-11 x AW141293 ..

Align seg 1/1 to: AW141293 from: 1 to: 504.

```

303 GluAsnGlyThrSerAsnLysAspGluGluPheArgMetProTyrLeu 319
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3 GAAATGCAACCCCTAATRAAGAGATGATTCACAGATGCCCTATCTAG 52
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
319 HisGlnGlnLeuProAlaGlyLeuProMetValProGluValAla 336
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
53 CCATCAGAGAGTGCAGCGGAGATTCTCCCATGTGTCAGAGGTTGCC 102
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
336 InAlaValGlyAlaSerGlnGlyHisHisThrLysAspPheThrArgAla 352
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
103 AGGCTGTGGGGTGTAGTCAGAGACACCAATCCAAAGATTTCGCCAGGCA 152
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
353 AlaProAsnProAlaLysAlaThrValThrAlaMetLeuAlaArgGlu 369
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
153 GCTCCAAATCTGCGCAAGGCCAGGTAAGTGCATGATGATGCGCAGAA 202
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
369 uLeuTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysSerAsnI 386
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

5'-GAGAGAGAAGACTCTCAGATCGCTTAATTAAATAATATACCCCCCCCCCCC-3'
and 3' adaptor sequence:
5'-GAGAGAGAGACTCCAGACTTTTTCCTTTTTTTT-3'. The library was
size-selected for >0.5 Kb inserts and has an average
insert size estimated at 1.2 kb. This library was
constructed using the CAP-trapper method for full-length
enrichment and has not undergone amplification. Library
was constructed by Dr. Claudio Schneider (LMCIB-Area
Science Park, Trieste, Italy). "
BASE COUNT      130 a      152 c      123 g      188 t      1 others
ORIGIN

alignment_scores:
    Quality:      836.00      Length:      178
    Ratio:        4.889       Gaps:      0
Percent Similarity: 96.067   Percent Identity: 89.888

alignment block:
US-09-316-048-11 x AM162002/rev ..

Align seg 1/1 to reverse of: AM162002 from: 1 to: 594

1 MetTyLysProValAspProHisSerArgMetGlnSerThrTyrSery 17
|||||.....:|||||||.....:|||||
534 ATGTATTAACCCTGTGGACCTTACTCTCGAAGCACTCCACCTATTAAC 485
```

[illegible]

```

463 GGCCAGGGGATCATATCGATTAGCCGACGTGGCCCAAGTCCAGACAGGCCAA 414
208 slysglulysglulproglutyrmetleuleuthrgluatrglyleuproa 225
413 AAAGAGAGAGAGCCACAGATACACGCTCTCCACAGAGACGAGCCTCCGCC 364
225 rgatgagatggluphevalmerclnvallysalglyshststpralaglu 241
363 GCCCGAGGAGCTTTGTGATCGAGGTGAAGGTGGAAACCCACACTGCAGAA 314
242 glyvalglythrassnlyslsvalalalysargasnalalaglulussme 258
313 GGAAAGGGGACACCAACAGAGAGTGGCCACACCGCATGCCACGAGAAACT 264
258 flaugluilleauglypheyvalproglinalagluproalalysproa 275
263 GCTGAGAGATCCTTGTTCAAAAGTCCCGCAGGCGCACACCCACCAACCCG 214
275 lalauylsersercluglulysthrprovalylslysprproglyaspglyatg 291
213 CACCCAAATCAGAGAGAGAAACACCCATAAAGAAACACGAGGGATGGAAGA 164
292 lysvalthrphephegluproserprogllyaspglussnlythrseras 308
163 AAAGTAACCTTTTGAACCTGCGTCTGGGGATGAAATGGACTAGTAA 114
308 nlyaspsplugluphearqmetprotyrleuserhisglinglualeuproa 325
113 TAAAGAGATGATTCAGAGATGCTTATCTAAGTACATCACCATCTCTC 64
325 laglylleu 328
63 CTGGACTTCTT 53
seq_name: gb_est75:BE781223
seq_documentation_block: 673 bp mRNA EST 20-SEP-2000
LOCUS BE781223
DEFINITION 601469159P1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3872247 5',
mRNA sequence.
ACCESSION BE781223
VERSION BE781223.1 GI:10202421
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 673)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LIC613 row: e column: 16
High quality sequence stop: 644.
Location/Qualifiers
1..673
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3872247"
/clone_lib="NIH_MGC_67"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-SpeK6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT

```

Average insert size 1.75 kb. Library constructed by Life Technologies.  
 BASE COUNT 241 a 156 c 178 g 98 t  
 ORIGIN

alignment\_scores:  
 Quality: 826.00 Length: 227  
 Ratio: 4.325 Gaps: 5  
 Percent Similarity: 84.141 Percent Identity: 76.652

alignment\_block:  
 US-09-316-048-11 x BE781223 ..

Align seg 1/1 to: BE781223 from: 1 to: 673

```

94 AsnLysSerGluIleSerGlnValPheGluIleAlaLeuLysArgAsnLeu 110
|||||
1 AATAAATCTGAATAATAGTCAAGCTTTGAGATTGCACTTAAAGGAACCTT 50
110 uproValAsnPhenGluValAlaArgGluSerGlyProProHisMetLysA 127
|||||
51 GCCTGCAATTTGAGAGTGCCCGGAGAGAGTGGCCACCCACCATGAAGA 100
127 snPheValThrArgValSerValGlyLupheValGlyLuglyLugly 143
|||||
101 ACTTGTGACCAAGGTTTCGGTGGGAGATTGTGGGGAGGTGAAGG 150
144 LysSerLysLysIleSerLysLysAsnAlaIleArgAlaValLeuGlu 160
|||||
151 AAAAGCAAGAGATTTCAAGAAATGCCGCATAGCTGTTCTTGAGGA 200
160 nueAArgArgLeuProLeuProLeuProAlaValGluArgValLysProArg 177
|||||
201 GCTGAAGAAAGTTACCCCGCTGCTCAGTTCAGACAGAGTAAGCCTAGA 250
177 leLysLysLysSerGlnProThrCysLys.....ThraIleProAspTyr 191
|||||
251 TCMAAAGAAACCAAAACCATATGTCAGCAGCAGCAGCAAGCCAGAAAT 300
192 GlyGlnGlyMetCAsnProIleSerArgLeuAlaGlnIleGlnGlnAla 208
|||||
301 GCGCAGGAGATCAATCCGATTACCGACAGTGGCCGATCCAGAGGCA 350
208 slySGluLysGluProGluTyrMetLeuLeuThrGluArgGlyLeuPro 225
|||||
351 AAAGGAAAGAGAGCCAGAGTACAGCTCTCACAGAGCGAGGCTCCCGC 400
225 rGArgArgGluPheValMetGlnValLysValGlyHisThrAlaGlu 241
|||||
401 GCGCAGGAGATTGTGATGAGAGTGAAGTTGGAACCACTGCAGAA 450
242 GlyValGlyThrAsnLysLysValAlaLysArgAsnAlaAlaGluAsn 258
|||||
451 GGAACGGGACCAACAAAGAGTGGCAAGCGCAATGCAAGCCAGAAAT 500
258 leuGluIleLeuGly..PheLysValProGlnAlaGlnProAlaLys... 273
|||||
501 GCTGGAGATCTTGGCTTCAAAAGTCCCGCAGGCGCAGCCCAACAAAC 550
274 .....ProAlaLeuLysSerGluLugly 281
|||||
551 CAAGACACCCCGCGAAATAAAATAAAGAAACAAATATAAATAAT 600
281 snThrProVal.....LysLysProGlyAspGlyArg 292
|||||
601 AACACAAATAAACACCCCAACATTAACAACAAAAAAGGGGGGGGGCC 650
292 yGValThrPhePheGluProSerProGly 301
|||||
651 AA.....AAAGTCCCGCGGG 667

```

seq\_name: gb\_est77:BE872137

seq\_documentation\_block:  
 LOCUS BE872137 852 bp mRNA EST 27-SEP-2000  
 DEFINITION 60146238P1 NIH\_MGC\_65 Homo sapiens cDNA clone IMAGE:3850496 5',  
 mRNA sequence.

ACCESSION BE872137 GI:10320913

VERSION BE872137.1  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 852)

AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://Image.lnl.gov>  
 Plate: LCM556 row: k column: 09

High quality sequence stop: 637.  
 Location/Qualifiers  
 1. 852  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3850496"  
 /cclone\_lib="NIH\_MGC\_65"  
 /tissue\_type="adenocarcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: colon; Vector: pCMV-Sport6; site\_1: NotI;  
 site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.8 kb. Library constructed by Life  
 Technologies."

FEATURES  
 source

BASE COUNT 275 a 215 c 218 g 144 t  
 ORIGIN

alignment\_scores:  
 Quality: 823.00 Length: 252  
 Ratio: 3.846 Gaps: 9  
 Percent Similarity: 84.921 Percent Identity: 73.810

alignment\_block:  
 US-09-316-048-11 x BE872137 ..

Align seg 1/1 to: BE872137 from: 1 to: 852

```

205 glnGlnAlaLysLysGluLysGluProGluTyrMetLeuLeuThrGlu 221
|||||
1 CAGCAGGCAAAAGAGAGAGGAGCAGAGTACACCTCTCAGCAGAGCG 50
221 gGlyLeuProArgArgArgGluPheValMetGlnValLysValGlyHis 238
|||||
51 AGGCTCTCCCGCGCGAGGAGTTGTGATGAGGTGAAGGTTGGAAAC 100
238 isThrAlaGluGlyValGlyThrAsnLysLysValAlaLysArgAsnAla 254
|||||
101 ACACCTGCAAGAGAGAGCGGCCACCAACAAGAGTGGCCAGCGCAATGA 150
255 AlaGluAsnMetLeuGluIleLeuGlyPheLysValProGlnAlaGln 271
|||||
151 GCGGAGAACATGCTGGAGATCTTGGTTCAAAAGTCCCGCAGGCGAGCG 200
271 oAlaLysProAlaLeuLysSerGluLuglyThrProValLysLysPro 288
|||||
201 CACCAAAACCGCAGCTCAAGTCAAGAGAGAAAGACACCCATMAAGAAC 250

```

288 LysAspGlyArg...LysValThrPhePheGluProSerProGlyAsp... 302  
 251 GGGATGGAACACACAGTCAACCTTTTTCAGCTGGCTGGGCGATGC 300  
 303 GluAsnGlyThrSer...AsnLysAspGluLysPheArgMetPro. 316  
 301 AACATGGCGCATCATTAATCACACAGCAGCATGATCCAGATGCCTT 350  
 317 ..TyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuProMetValPro 332  
 351 CATCTAAGCTCATTCACAGCAGCTGGCTGGTAATCTTCCATGAGTCC 400  
 333 GluValAlaGlnAlaValAlaGlyValSerGlnGlyHisHisThrLysAsp 349  
 401 GAGTCCGCCAGCGCTGAGAGATTAGTCAAGAGCATCACACCAAGATT 450  
 349 ethArg...AlaAlaProAsnProAlaLysAlaThrValThrAlaMet. 364  
 451 TACCAGGAGACAGCTCCGACATCTGCACAAAGCCAGCTAATCAGCTGC 500  
 365 .IleAlaArgGluLeuLeuTyrGlyGlyThrSerProThrAlaGluThr 381  
 501 ACTAGCCCGAGAGTCTGATGAGGGGCGCCCGCCACAGCCGAGACCA 550  
 381 LeuLysSerAsnIleSerSerGlyHisValProHisGlyProArgThr 397  
 551 TTTTACAGATTAACATCTCTTCAGCGCCAGCATGACCATGACCTCAGC 600  
 398 ArgProSerGlnGlnLeuTyrTyrLeuSerArgAlaGlnGlyPheGln.V 414  
 601 AGACCTCTGAGCAACTGACTATCTACCAAGAGTCCAGCATACCAAG 650  
 414 aLgUuTyrLysAspPheProLysAsnLysAsnGlyLysValSer.Le 430  
 651 TTGACTACAAAGACCTTCCCAAAAGCAAAAGCAAAAGTATCTCTT 700  
 430 uLleAsnGlySerSerGlnProProLeuValSerHisGlyIleGlyLys 446  
 701 AATCAATTGGTACTCTCAGCAACGCTGTGATCAGAGAAATCGGGAAG 749

seq\_name: gb\_est76:BE810330

seq\_documentation\_block:

LOCUS BE810330 609 bp mRNA EST 21-SEP-2000  
 DEFINITION MR4:PT0052-220500-206-f06 PT0052 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BE810330  
 VERSION BE810330.1 GI:10242524

KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 609)  
 Dias Neto,E., Garcia Correa,R., Verjowski-Almeida,S., Briones,M.R.,  
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
 Brunstein,A., deoliveira,P.S., Bucher,P., Jongehele,C.V., O'Hare  
 ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
 Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 MEDLINE 20202663

COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil

Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=tl2-MR4-PT0052-220  
 500-206-f06&tl=2000-05-22&tl=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 10  
 High quality sequence stop: 589.  
 Location/Qualifiers

#### FEATURES

source  
 1..609  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="PT0052"  
 /dev\_stage="Adult"  
 /note="Organ: pnet; Vector: puc18; Site\_1: Sma1; Site\_2:  
 Sma1; A mini-library was made by cloning products derived  
 from ORESTES PCR (U.S. Letters Patent application No. 196  
 ,716 - Ludwig Institute for Cancer Research) profiles  
 into the pUC 18 vector. Reverse transcription of tissue  
 mRNA and cDNA amplification were performed under low  
 stringency conditions."  
 BASE COUNT 173 a 153 c 145 g 138 t  
 ORIGIN

alignment\_scores:  
 Quality: 808.50 Length: 185  
 Ratio: 4.673 Gaps: 2  
 Percent Similarity: 93.514 Percent Identity: 89.189

alignment\_block:

US-09-316-048-11 x BE810330 ..  
 Align seg 1/1 to: BE810330 from: 1 to: 609

304 AsnGlyThrSerAsnLysAspGlnGlu.PheArgMetProTyrLeuSerH 320  
 8 CATGGAGCTAGTAATAAG...AGCATGAGTCCAGATGCTTATCTAAGTC 54  
 320 IsGlnGlnLeuProAlaGlyIleLeuProMetValProGluValAlaGln 336  
 55 ATCAGACAGCTCCCTGCTGA.ATTCTTCCATGCTGCCGAGGTGCCAG 103  
 104 GCTGTAGAGATTAGTCAAGAGCATCACACCAAGATTATTTACAGGCGAC 153  
 337 AlaValAlaGlyValSerGlnGlyHisHisThrLysAspPheThrArgAla 353  
 104 GCTGTAGAGATTAGTCAAGAGCATCACACCAAGATTATTTACAGGCGAC 153  
 353 apProAsnProAlaLysAlaThrValThrAlaMetIleAlaArgGluLeu 370  
 154 TCCGAATCTCTGCCAGGCGCACGCTAAGTGCATGATAGCCGAGAGTAGT 203  
 370 euTyrGlyGlyThrSerProThrAlaGluThrIleLeuLysSerAsnIle 386  
 204 GGTATGGGGGACCTCGCCACAGCCGAGACCATTTTAAACAATAACATC 253  
 387 SerSerGlnHisValProHisGlyProArgThrArgProSerGlnGln 403  
 254 TCTTCAGGCGCCAGTACCCATGAGACCTCTCAGAGACCTCTGAGCAATC 303  
 403 uTyrTyrLeuSerArgAlaGlnGlyPheGlnValAlaGlyLysAspPhe 420  
 304 GGAATATCTTCCAGAGTCCAGAGTCCAGGATTCACAGTTCAATCAAAACAT 353  
 420 roLysAsnAsnLysAsnGlyLysValSerLeuIleAsnGlySerSerGln 436  
 354 CCAAAACACACAGAACGAATTTGATCTTTATCAATAGGCTCCCTCAG 403  
 437 ProProLeuValSerHisGlyIleGlyLysAspValGluSerCysHisAs 453  
 404 CCACCTGTGATCAGCAGCATGATCGGCAAGGATGGAGTCTGCGCAAGA 453  
 453 pMetAlaLysLeuAsnIleLeuLysLeuLeuSerGluLeuAspGlnGln 470  
 454 TATGGCTCGCTGAACATCTTAAGTTCTGTGAGTAGAGCAACACAA 503

470 eethrclumetProArGThrGlyAsnGlyProValSerAlaGlyArg 486  
 |||||||  
 504 GTACAGAGATGCCAAGACGAAAGCAACCATGTCTGTGTGGAGG 553

487 Cys 487  
 |||  
 554 TGC 556

seq\_name: gb\_est40:AW163206

seq\_documentation\_block:

LOCUS AW163206 564 bp mRNA EST 09-NOV-1999  
 DEFINITION au97f05.x1 Schneider fetal brain 00004 Homo sapiens cDNA clone  
 IMAGE:2784225 3' similar to TR:095793 095793 STRAUPEN PROTEIN. ;  
 mRNA sequence.

ACCESSION AW163206  
 VERSION AW163206.1 GI:6302239  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 564)  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
 Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,  
 J., Moore,B., Scheinberg,K., Stepcoe,M., Tan,F., Theisling,B.,  
 White,Y., Wylie,T., Waterston,R. and Wilson,R.

Washu-NCI human EST Project  
 Unpublished (1997)

TITLE

JOURNAL  
 COMMENT Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800  
 Fax: 314 286 1810

Email: est@wustl.edu

This clone is available royalty-free through LNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -40UP from Gibco

High quality sequence stop: 455.  
 Location/Qualifiers

FEATURES

SOURCE

1..564

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2784225"

/clone\_lib="Schneider fetal brain 00004"

/sex="male"

/tissue\_type="frontal lobe"

/dev\_stage="5 months post-conception"

/lab\_host="DH10B"

/note="Organ: brain; Vector: pBluescript SK (Stratagene);  
 Site\_1: SstI; Site\_2: XhoI; Double-stranded cDNA was  
 prepared from human fetal brain tissue. 5' and 3'  
 adaptors were used in cloning as follows: 5' adaptor  
 sequence:  
 5'-GAGAGAGAGAGAGAGCTCAAGATCTTAATTAAATTATCCCCCCCCC-3'  
 and 3' adaptor sequence:  
 5'-GAGAGAGAGAGAGCTCAAGATCTTAATTAAATTATCCCCCCCCC-3'  
 size-selected for >0.5 kb inserts and has an average  
 insert size estimated at 1.2 kb. This library was  
 constructed using the CAP-trapper method for full-length  
 enrichment and has not undergone amplification. Library  
 was constructed by Dr. Claudio Schneider (LNCIB-Area  
 Science Park, Trieste, Italy)."

BASE COUNT 125 a 147 c 117 g 175 t

ORIGIN

alignment\_scores:

Quality: 808.00

Ratio: 4.838

Percent Similarity: 93.820

alignment\_block:

Length: 178

Gaps: 0

Percent Identity: 88.202

US-09-316-048-11 x AW163206/rev ..

Align seg 1/1 to reverse of: AW163206 from: 1 to: 564

1 MetTyrLysProValAspProHisSerArgMetGlnSerThyrSerTy 17

||||| ||||| .....

534 ATGTATTAGCGCTGTGACCCCTTACTCTCGATGTCACCTCCACTATAC 485

17 rgLysMetArgGlyAlaTyrProProArgTyrPheTyrProPhePro 34

||||| ||||| .....

484 CCACATGAGAGAGAGGTCTTATCCCGAGAGTACTTTACCAATTCAG 435

34 aLProProLeuLeuTyrGlnValAlaGluLeuSerValGlyGlnGlnPhe 50

||||| ||||| .....

434 TTCACCTTACTTATCAAGTGAAGTCTCTGTGGAGAGACGAATTT 385

51 AsnGlyLysGlyLysMetArgProProValLysHisAspAlaProAla 67

||||| ||||| .....

384 AATGCCAAGGAAAGACACAGAGGCTGCGAAGACAGATGCTGTCACA 335

67 gAlaLeuArgTyrThrLeuGlnSerGlnProLeuProGluArgGluVal 84

||||| ||||| .....

334 AGCGTTGAGAGATCTTCAGATATAGCCCTGCGAGAGAGGCTGAGAG 285

84 snGlyArgGluAlaGluGluGluAsnLeuAsnLysSerGlnLysSerGln 100

||||| ||||| .....

284 ATGGAAGAGATCGAAGAAAGAAATCTCAATTAATCTGAATTAAGTCAA 235

101 ValPheGluIleAlaLeuLysArgAsnLeuProValAsnPheGluVal 117

||||| ||||| .....

234 GTGTTTGAGATTGCACTTAACGGAAGTCCCTGTAATTTGAGGTGGC 185

117 aArgGluSerGlyProProHisMetLysAsnPheValThrArgValSerV 134

||||| ||||| .....

184 CCGGAGAGTGGCCACCCACATGAGACTTTGACCAAGCTTTCGG 135

134 aLgLyLupPheValGlyGluGlyGlyLysSerLysLysLysLys 150

||||| ||||| .....

134 TTGGAGAGTTTGTGGGGAAGTGAAGGAAAGCAAGAAAGATTTCAG 85

151 LysAsnAlaAlaArgAlaValLeuGlnLeuArgLysArgLeuProLeu 167

||||| ||||| .....

84 AAAAATGCCCGCATGACTCTTCTTGAGAGAGCTGAAGATGACGCCCT 35

167 upProAlaValAlaArgValLysProArgLys 178

||||| ||||| .....

34 GCCTGCAGTTGAACGAGTAAAGCTTAGAAAAA 1

seq\_name: gb\_est73:BE617223

seq\_documentation\_block:

LOCUS BE617223 749 bp mRNA EST 24-AUG-2000  
 DEFINITION 601441906r1 NIH\_MGC\_65 Homo sapiens cDNA clone IMAGE:3846265 5',  
 mRNA sequence.

ACCESSION BE617223  
 VERSION BE617223.1 GI:9888161  
 KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 749)  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.

National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be







```
280 GluLysThrProValLysLysProGlyAspGlyArgLysValThrPhePhe 296
|||||
401 GAGAAAGACACCCATAAAGAACCGAGGATGGAAGAAAGTAACCTTTT 450
|||||
296 eGIuProSerProGlyAspGluAsnGlyThrSerAsnLysAspGluGluP 313
|||||
451 TGAACCTGGCTCTGGGATGAACTAGGACTAGTACTAAAGAGATGAGT 500
|||||
313 heArgMetProTyrLeuSerHisGlnGlnLeuProAlaGlyIleLeuPro 329
|||||
501 TCAGGATGCTTATCTGTCATCAGCAGCGCTGCTGGAATCTTCCC 550
|||||
330 MetValProGluValAlaGlnAlaValGlyValSerGlnGlyHisHisTh 346
|||||
551 ATGGTGACCGACGCTCCG.CAGGCTGAGAGTCAGTCAGGACATCACAC 599
|||||
346 rLysAspPheThrArgAlaAlaProAsnProAlaLysAlaThrValThra 363
|||||
600 CACAGATTAAACAGCCAGCATCCGAATCCTGCCAAGAGACCACCGTACTG 648
|||||
363 IamEtIleAlaArgGluLeuLeuTyrGlyGlyThrSerProThrAlaGlu 379
|||||
649 CACAAAGATAGCCCGAGAGTGGGAGGAGGCGACCTCGACCA..... 689
|||||
380 ThrIleLeuLysSerAsnIleSerSerGlyHisValProHisGly..... 394
|||||
690 .....CAGCGAAACATATACACAGCAACACCTCTCTCAGGCGCAAGTA 730
|||||
395 .ProArg 396
|||||
731 ACCCAGG 737
```

